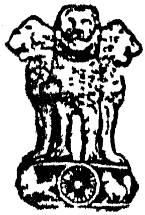


सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI

I.A.R.I.6.
GIP NLK—H-3 I.A.R.I.—10-5 55—15,000

PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL 196

LONDON
Published for the Royal Society by the
Cambridge University Press
Bentley House, N.W.1

22 April 1949

Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)
and published by the Cambridge University Press
Cambridge, and Bentley House, London

Agents for Canada and India: Macmillan

CONTENTS

SERIES A VOLUME 196

No. A 1044. 22 February 1949

PAGE

Presidential Address. By Sir Robert Robinson v

Flow of a compressible fluid about a cylinder. II. Flow with circulation. By T. M. Cherry 1

Numerical solutions for transonic flow. By T. M. Cherry 32

Order-disorder statistics. I. By C. Domb 36

Neutron diffraction. By G. E. Bacon and J. Thewlis 50

Some crystal-boundary phenomena in metals. By B. Chalmers. (Plate 1) . . 64

A general kinetic theory of liquids. VI. The equation of state. By A. E. Rodriguez 73

Some theoretical considerations on the dynamic properties of plastics. By N. F. Astbury 92

The influence of diffusion on flame propagation. By A. G. Gaydon and H. G. Wolfhard 105

The second virial coefficients of organic vapours. By J. D. Lambert, G. A. H. Roberts, J. S. Rowlinson and V. J. Wilkinson 113

The structure and properties of the alloy Cu_2MnIn. By B. R. Coles, W. Hume-Rothery, F.R.S. and H. P. Myers 125

No. A 1045. 22 March 1949

Non-central interactions between neutron and proton. By Tsi-Ming Hu and H. S. W. Massey, F.R.S. 135

Influence of chlorine substitution on the rate of oxidation of paraffin hydrocarbons. By C. F. Cullis, Sir Cyril Hinshelwood, F.R.S. and M. F. R. Mulcahy . . 160

A dynamical model of a crystal structure. II. By Sir Lawrence Bragg, F.R.S. and W. M. Lomer. (Plates 2 and 3) 171

A dynamical model of a crystal structure. III. By W. M. Lomer . . . 182

Vibration spectra of hydrocarbon molecules. I. Frequencies due to deformation vibrations of hydrogen atoms attached to a double bond. By N. Sheppard and G. B. B. M. Sutherland 195

The intensity distribution in the nitrogen band systems emitted from the earth's upper atmosphere. By D. R. Bates 217

The indefinite metric in relativistic quantum mechanics. By K. L. Le Couteur . 251

The Gouy diffusiometer; further calibration. By A. G. Ogston 272

A systematic treatment of moving axes in hydrodynamics. By G. C. McVittie . 285

Two-dimensional fluid motion referred to a network of orthogonal curves. By G. C. McVittie 301

No. A 1046. 7 April 1949

PAGE

Bakerian Lecture. Nuclear explosions. By Sir George Thomson, F.R.S. (Plates 4 to 7) 311

Nuclear disintegrations produced by cosmic rays. By J. B. Harding, S. Lattimore and D. H. Perkins 325

A resonating-valence-bond theory of metals and intermetallic compounds. By L. Pauling, For.Mem.R.S. 343

The polymerization of acetaldehyde at low temperatures. By J. C. Bevington and R. G. W. Norrish, F.R.S. 363

Splashes from underwater explosions. By H. Kolsky, J. P. Lewis, M. T. Sampson, A. C. Shearman and C. I. Snow. (Plates 8 to 11) 379

Experimental and theoretical studies of the behaviour of slow electrons in air. I. By L. G. H. Huxley and A. A. Zaazou 402

Experimental and theoretical studies of the behaviour of slow electrons in air. II. Ionospheric and other applications. By L. G. H. Huxley 427

Preparation and properties of a reproducible neutron standard. By E. Bretscher, G. B. Cook, G. R. Martin and D. H. Wilkinson 436

No. A 1047. 22 April 1949

The kinetics of the interaction of atomic hydrogen with olefines. I. Apparatus and use of para-hydrogen techniques. By H. W. Melville, F.R.S. and J. C. Robb 445

The kinetics of the interaction of atomic hydrogen with olefines. II. Diffusion theory. By H. W. Melville, F.R.S. and J. C. Robb 466

The kinetics of the interaction of atomic hydrogen with olefines. III. Theory and technique involved in the use of the oxides of molybdenum and tungsten as hydrogen atom removers. By H. W. Melville, F.R.S. and J. C. Robb . . 479

The kinetics of the interaction of atomic hydrogen with olefines. IV. Results obtained by the foregoing techniques. By H. W. Melville, F.R.S. and J. C. Robb 494

Molecular orbitals and the Hartree field. By W. Moffitt 510

The electronic structures of carbon monoxide and carbon dioxide. By W. Moffitt 524

The kinetics of certain reactions between methyl halides and anions in water. By E. A. Moelwyn-Hughes 540

The spectral response of a phosphor to infra-red radiation. By E. F. Daly. (Plate 12) 554

The emission of the negative system of nitrogen from the upper atmosphere and the significance of the twilight flash in the theory of the ionosphere. By D. R. Bates 562

Index 592

Address of the President
Sir Robert Robinson, at the
Anniversary Meeting, 30 November 1948

On behalf of the Fellows of the Royal Society the following telegram was sent on 16 November to His Majesty the King:

> 'We your loyal and dutiful subjects the President, Council and Fellows of the Royal Society of London beg leave to offer our respectful and sincere felicitations on the birth of the Prince.'

A message was also sent to Her Royal Highness The Princess Elizabeth as follows:

> 'The President, Council and Fellows of the Royal Society of London send their respectful and sincere congratulations to Your Royal Highness on the happy event of the birth of the Prince.'

We have received gracious acknowledgements in the following terms:

> 'Please convey to the Council and Fellows of the Royal Society the sincere thanks of the Queen and myself for their kind message on the birth of our grandchild.
>
> GEORGE R.'

> 'We are most grateful for your kind message of congratulations.'
>
> ELIZABETH AND PHILIP.'

Award of Medals, 1948

The COPLEY MEDAL is awarded to Professor ARCHIBALD VIVIAN HILL, C.H., for his outstanding contributions to the physiology of muscular processes.

Towards the end of the nineteenth century physiologists discerned a field of investigation of the physical and chemical phenomena involved in the characteristic activity of muscle and nerve, but the questions propounded could receive no satisfactory answers, chiefly owing to the lack of adequate methods. Hill, during the last thirty-five years, has applied profound mathematical skill and a flair for the design and management of delicate instruments to the provision of a sound basis for experimental and theoretical advance. In his hands the thermopile was developed into an instrument of great precision and delicacy, and conditions for its use in relation to the material studied were established so as to provide significant results.

The work on heat production in muscle has led to a picture within a framework which can accommodate many contemporary biochemical studies in a harmonious fashion. Attention has constantly been paid to ensuring the thermodynamic soundness of the systems suggested. Moreover, he was able to exploit some of his findings on isolated muscle in the course of a study of the dynamics of muscular contraction in the living body, including that in the human subject.

A natural extension of Hill's researches has enabled him to contribute notably to our knowledge of the heat production and excitability of nerve. Previous workers failed to detect any heat production in nerve tissues, but Hill demonstrated the definite thermal changes that are associated with the passage of impulses. These were found to be so small that no one-way contemporary chemical reaction could account for them. It would be necessary to postulate reversible changes which proceed in both directions. This was a main factor in the elaboration of the modern view that the nerve impulse results in an altered state of polarization of the essential surface membranes. The heat change which occurs after the passage of the impulses was found to be small but definite, and was related to the energy expended in the recovery of the membrane.

Hill has made contributions of great distinction in other fields, and among other topics investigated by him is that of the physical chemistry of the oxyhaemoglobin equilibrium in the blood.

Hill has been a source of inspiration to many pupils, collaborators and colleagues and has founded a school of research. His eminent services to physiology have been possible as the result of a happy combination of wide basic knowledge of the physical sciences, unusual skill in the design and execution of experiments, a philosophic outlook, and above all the curiosity and ambition to attack some of the most difficult problems in biophysics and biochemistry.

The RUMFORD MEDAL is awarded to Professor FRANZ EUGEN SIMON, C.B.E., for his distinguished researches on the properties of matter at low temperatures.

The school of low-temperature physics which Simon has created at Oxford is one of the leading centres of such research in the world and has achieved highly important results in the spheres of both theory and practice.

His great knowledge of thermodynamics was brought into service in the development of extremely simple methods for reaching the temperatures of liquid helium in the laboratory. He realized that at low temperatures adiabatic expansion of helium compressed into a small reservoir would absorb sufficient heat to liquefy the bulk of the gas and to cool the container and specimens under examination to liquid helium temperatures. Thus, provided reasonable supplies of liquid air are available, it becomes possible for any laboratory to carry out investigations at the lowest attainable temperatures with equipment which is inexpensive and simple to operate. Many other contributions to technique in this field have been made by Simon and his co-workers.

With the aid of these new or improved methods many interesting observations, with a bearing on fundamental knowledge, have been made. Outstanding problems in superconductivity, especially the role played by impurities, have been solved, and the 'freezing out' of magnetic flux has been explained. The peculiar properties of liquid helium II and the transition to this state have been investigated intensively. The thermodynamics of the approach to absolute zero has been stated very clearly by Simon and progress has been made in new methods for reaching this temperature. More recently, extended equipment has facilitated the study of the properties of atomic nuclei oriented in magnetic fields.

Simon has made significant contributions to the elaboration of devices for the separation of isotopes and to the design of plants in which thermodynamic considerations are of predominating importance.

He has aptly illustrated the thesis that, in the world of modern physics, discovery and the progress of theory follow surely on the introduction of new and powerful tools of research.

A ROYAL MEDAL is awarded to Professor HAROLD JEFFREYS for his fundamental investigations in theoretical geomechanics.

Much of Jeffreys's well-known book *The Earth* (1924, 1929) is based on his own work, and this is especially true of the section on near earthquakes. However, the book owes its unique position not only to these original contributions but also to the skill and insight shown in the building up of a self-consistent picture of the earth from the very diverse and apparently contradictory materials available at the time when it was written. Its great influence has been due to the width of interest and the critical skill as well as to the depth of learning of its author.

During the thirties Jeffreys's main interest was in seismology. In a series of papers he has discussed the available observations in detail and reduced the errors in the seismological time-tables by a factor of 10 or more. This work has had an important effect on our knowledge of the earth's interior, and has yielded numerical information of surprising accuracy, and, for example, the radius of the earth's central core is now known with an accuracy of a few kilometres.

Some of his early papers and the book *Scientific Inference* show an interest in the philosophical basis of the scientific method and in the theory of errors. This was intensified by the difficulties he met in adequately treating the seismological travel times, to which orthodox 'least square' methods are inapplicable. The discussion of these observations led to a series of papers on significance tests, and, finally, to a book, *The Theory of Probability*, in which he discusses *de novo* the whole problem of the relation between observation and 'physical quantities'. The departures from orthodoxy have not proved universally acceptable, but it is generally agreed that Jeffreys has made an important contribution to a subject on which the last word has not yet been said.

Other topics to which Jeffreys has made material contributions are pure mathematics, the origin of the moon, the constitution of the planets, hydro- and aerodynamics. In particular, he gave the first mathematical discussion of the theory of aerofoils of finite thickness.

Other less relevant examples of the catholic interests of the Plumian Professor of Astronomy and Natural Philosophy could be cited. Whether as a specialized applied mathematician or as a teacher and leader of research, the eminent services he has rendered are most widely recognized.

A ROYAL MEDAL is awarded to Professor JAMES GRAY, C.B.E., for his distinguished work on the mechanism of posture and locomotion in vertebrate and other animals.

In 1933 Gray, by the use of cinematography, analyzed the forces involved in the progression and turning of fish with long and short bodies. The inquiry reached its full development during the last ten years with the extension of these studies to vertebrates generally. In 1936 he discussed the swimming of dolphins, raising the hydrodynamical problem of the type of waterflow past an undulating body. Later he analyzed the locomotion of Amphibia and of snakes. The results of this part of the work led to a discussion of the mechanics of the tetrapod skeleton which was fundamental and placed an important part of the comparative anatomy of vertebrate animals on a functional basis for the first time. His work shows that the whole body of a tetrapod must be considered as a single functional unit, changes in the tension of one muscle being accompanied by a demonstrable pattern of changes in other muscles. Gray analyzed and described the mechanical problems solved by tetrapods standing and moving over ground of varying slope and smoothness; he has also analyzed the neuro-muscular phenomena responsible for the evident co-ordination of action. He showed the importance of peripheral reflexes in the initiation and maintenance of swimming movements in fish. This led him to the problem of the origin of locomotory patterns. His experiments on the neuro-muscular system of annelids showed the major importance of patterns of peripheral stimuli.

This work, summarized in a Croonian Lecture, was later greatly extended by demonstration of the importance of peripheral reflexes in the locomotor activities of toads. It was found that there was no positive evidence for the existence of central rhythm in these animals, while, on the other hand, the proprioceptive reflexes arising from the muscles provide an adequate basis for locomotor action.

Gray's work is responsible for an important change in the direction of research in the comparative anatomy of vertebrates. Before its appearance skeletal structure was examined in extraordinary detail, and yet little or no attention was paid to musculature and to the relation of structure to the whole system in action. More than any other, Gray has shown that comparative anatomy is certainly not a dead science.

The DAVY MEDAL is awarded to Professor EDMUND LANGLEY HIRST for his distinguished contributions to the chemistry of the carbohydrates.

The first demonstration of the six-membered ring structure of an aldose derivative was offered by Hirst & Purves in their study of methylxyloside which was published in 1923. Since then, in collaboration with Sir Norman Haworth over a number of years and independently, Hirst played a great part in the establishment of fundamentals in the field of carbohydrate structure. His prolific experimentation is characterized by the highest accuracy and has afforded conclusive answers to many questions of importance.

Of his joint work in the Birmingham School the following must be cited: the recognition of pyranose and furanose forms of the simple sugars, the isolation of a crystalline furanoside for the first time, and the determination of the constitution of sucrose and other disaccharides.

His independent work on optical rotatory dispersion in the carbohydrate derivatives introduced new methods of attack of the constitutional problems which were of immense service in his later work on the constitution of vitamin C. In that investigation of outstanding merit he determined the structure of ascorbic acid and studied a large number of related products. In 1933 he shared in the synthesis of ascorbic acid and in the study of the physiological activity of the synthetic vitamin.

His researches on the polysaccharides have been conspicuously successful over a wide field, largely because of his development of new methods of end-group assay in the determination of structure. He has applied these methods successfully to starch, glycogen and other polysaccharides, and to him we owe many advances in our knowledge of the nature of alginic acid, the plant gums and pectic substances.

The regularities and irregularities that have been disclosed have been of equal interest and unexpectedness.

Hirst has been a pioneer in unravelling the complexities of carbohydrate architecture, and to him must be attributed much of our present knowledge of this group of such great biological significance.

The DARWIN MEDAL is awarded to Professor RONALD AYLMER FISHER for his distinguished contributions to the study of biological evolution.

A general principle that is consistently developed in Professor Fisher's writings is that the course of evolution is not controlled by mutation but by selection operating upon the heritable variability which Mendelian recombination supplies.

In 1928 he was the first to suggest that dominance is the product of selection operating in the gene complex of the organism. This theory he tested and confirmed experimentally with domestic poultry, and later with mice.

The implications of his concept extend beyond the dominance phenomena, and underlie the whole theory of the modification of genic effects through selection acting on the gene complex. This enables a character, even when under unifactorial control, to be adjusted to the needs of the organism. To Darwin, selection was a mechanism for rejecting the bad and conserving the good; it is largely to Fisher that we owe the view that it may mitigate the bad and enhance the good effects of the hereditary units.

The studies of Fisher on abundance and variability, the first of them undertaken in collaboration, have provided means for proving in general terms the proposition, implicit in the modern concept of evolution, that genes having minute favourable effects are in process of being spread through populations and are bringing about evolutionary change at the present time.

Fisher's methods of population analysis, developed from 1930 onwards, have made it possible to study the numerical aspect of animal communities with an exactitude previously unattainable.

The special opportunities for evolutionary study presented by polymorphism have always attracted Fisher, and he has repeatedly analyzed the balance of selective agencies which may be involved. This point of view is apparent in his extensive work on human serology. He has also studied a polymorphism which had especially interested Darwin, that of the heterostyled plants, and he has extended the subject to elaborate the general theory of crossing-over in polyploids: a pioneer work of wide application.

Fisher's great contributions to statistics and to experimental design have provided tools now deemed essential to the quantitative biologist. But they are fundamental also to the experimental studies of evolution which have become so important a feature of modern biological research.

The HUGHES MEDAL is awarded to SIR ROBERT WATSON-WATT, C.B., for his pioneer researches in radiotelegraphy.

Watson-Watt has been a leader in the field of radio research since the early nineteen-twenties. He began to work on the subject of atmospherics in 1915, paying attention at first to the subject of the direction of arrival. He published the first English studies on this subject, establishing the 'cum-solar' swing of the place of origin throughout the day. He developed the cathode-ray direction-finder as a means for finding the direction of arrival of individual atmospherics. This proved of immense value in the location of thunderstorms during the war. With a number of collaborators he later studied the wave form of atmospherics.

He has been the leading exponent of the use of the cathode-ray oscillograph for a great variety of purposes in physical research.

Watson-Watt was the leader of the earliest British work on radar, or radiolocation as it was then called. Although the basic principles were known, enormous technical difficulties had to be overcome before radar could be developed in the form of an operational instrument. That these difficulties were in fact overcome was due, in this country, more to Sir Robert Watson-Watt than to any other man. He is now turning with equal insistence to the peacetime applications of radar, especially in the service of civil aviation.

I wish to thank Fellows of the Society and members of Council for assistance in preparing notes on the achievements of the Medallists.

Sir Alfred Egerton now lays down his office as Physical Secretary of the Society. During the past ten years he has shown a deep concern for the Society's welfare and the utmost willingness to undertake onerous duties.

I would like to remind you of some of his many activities. In the early days of 1939 he was chiefly instrumental in the creation of a Central Register of Scientists. The significance of this action by the Society was by no means exhausted by the cessation of hostilities. Almost in the first month of the war he took a leading part in persuading the War Cabinet to set up a Scientific Advisory Committee. These efforts, after initial setbacks, succeeded in October 1940, and the Committee functioned usefully during the war period and was the forerunner of the present Advisory Committee on Scientific Policy. I cannot deal at length with his services in a more personal capacity, for example, to the Ministry of Fuel and Power, but we recall with gratitude his interest in the reorganization of the Gassiot Committee and in the survey, which the Society conducted, of the needs of fundamental science after the war. He has been an ever-welcome guest overseas, and has a distinguished record as an ambassador of science. His great services to the Empire Scientific Conference of 1946 and its offspring, the Scientific Information Conference of the present year, are fresh in our memory. He worked zealously, carefully and effectively—an unusual and difficult combination—and the results will benefit every part of the Commonwealth in which scientific activity exists.

Sir Alfred will continue to act as virtual representative of the Society in several important capacities; if I may mention just one as an example, it may be as a member of the Scientific Advisory Council for the Festival of Britain 1951.

We shall not be able to take so much advantage of his habitual generosity as in the past, but, in case of need, we shall certainly seek his help and advice.

You have just elected Professor David Brunt as the Physical Secretary of the Society. To him we offer the most cordial welcome and good wishes coupled with thanks for his willingness to assume the duties and responsibilities of the office.

One of the most valuable of the traditions of the Society is our claim to exercise initiative when the circumstances suggest a new departure. So far as I am aware it has not previously been thought desirable to designate any person as Honorary Librarian to the Society, but the Council felt that the present Chairman of the Library Committee had given services which were probably without precedent. Professor Andrade's knowledge of the older scientific literature is unrivalled and has been freely available to the great advantage of the Society.

The retirement of Mr H. W. Robinson as Librarian occurred at the end of March of this year. He joined the Staff in 1902 and assisted the chief clerk and the Librarian; in 1910 he became Library Assistant and in 1930 Assistant Librarian. He has been Librarian since 1935. His work for the Society and its Library has been invaluable, and we may recall with special gratitude the care he showed in the safe bestowal of our treasures in the time of danger. An earnest student of the history of science, he has paid special attention to the study of Robert Hooke on whose life and work he is a recognized authority. In expressing our gratitude for his devoted labours, we note with pleasure that Mr Robinson will continue to assist Professor Turnbull in editing the Newton Letters.

The outstanding event of the past year in connexion with the Society has been the Scientific Information Conference. We welcomed delegates from the countries of the Commonwealth and also from the National Academy of Sciences of the United States, and from U.N.E.S.C.O. It can thus be claimed that the whole of the English-speaking scientific world was well represented. More than half of all scientific literature is written in English, and in making such a comprehensive review of scientific information services, I think the Society may be judged to have taken a timely initiative. We were gratified by the interest shown by Fellows and the time and energy so freely given by them to ensure the success of the Conference.

Your Council received from the Conference many recommendations dealing with the publication of papers reporting original work, abstracting services, reviews and annual reports, library services, classification and indexing, and other relevant matters. It has been decided to implement these, as far as proves possible, with the aid of a Treasury grant, and for this purpose the Council has set up an Information Services Committee which has just held its first meeting. I believe that an energetic follow-up of the recommendations will result in real progress. The value of the Conference itself, irrespective of these recommendations, should not be underestimated. Divergent points of view were advocated, and many of the difficulties besetting the free and rapid communication of new knowledge were at least recognized as the result of discussion and, in many cases, were removed. The problem is a complex one, and its solution demands both due regard for traditional methods and an acute appreciation of novel techniques.

Having given this lead the Society will now be expected to play an increasingly important part, nationally and internationally, in improving scientific information services.

It is doubtful whether the general progress of science has ever been more impressive than at the present time, and this fact emphasizes the importance of adequate outlets for the publication of the great volume of original work that is being poured out. Formidable archives are undoubtedly accumulating. To meet the demand many new journals have been founded and here is a short list of them. *Agricultural Chemicals*, Baltimore, 1946; *British Journal of Pharmacology and Chemotherapy*, U.K., 1946; *Hungarica Acta Chimica*, Budapest, 1946; *Journal of Colloid Science*, N.Y., 1946; *Journal of Polymer Science*, N.Y., 1946; *Zeitschrift für Naturforschung*, Wiesbaden, 1946; *Acta Chemica Scandinavica*, 1947, Chem. Soc. Denmark, Finland, Norway and Sweden; *Analytica Chemica Acta*, N.Y., 1947; *Biochemica et Biophysica Acta*, N.Y. 1947; *Die Makromolekulare Chemie*, Basle, 1947; *Food Technology*, N.Y., 1947; *Heredity*, London, 1947; *Journal of Glaciology*, U.K., 1947; *Nucleonics*, N.Y., 1947; *Acta Crystallographica*, U.K., 1948; *Acta Physica Austriaca*, Vienna, 1948; *Annali di Geofisica*, Rome, 1948; *Australian Journal of Scientific Research*, Australia, 1948; *Deutsche Hydrographische Zeitschrift*, Hamburg, 1948. Furthermore, the Scientific Reviews are constantly being extended in scope; one can even be thankful that they show a tendency to overlap. *Nature*, playing a more important part than ever before, is now supplemented by *Research*, whilst *Endeavour*, *Science Progress* and *Discovery* worthily maintain our reputation in the field of popular exposition. The press is far better informed on scientific affairs than was the case a few years ago, and we can anticipate useful autocatalysis in the reaction between public appreciation and public information.

A year ago I mentioned the formation of a committee to study the means whereby adequate accommodation for the scientific societies may be made available, and tactful reference to its activities will be found in the Report of Council. Without being too indiscreet it is possible to provide some further information on a subject of great interest to the Fellows of the Society. In the first place the Scientific Accommodation Committee has so far considered only the long-term problem, and it is matter for congratulation that representatives of so many interested parties reached full agreement on this aspect. It was unanimously agreed that the institution of a Science Centre would provide the best solution. Naturally everybody would like to know where that will be and exactly what the proposal involves. In regard to the first point a specific suggestion is being explored, and we hope that it will soon be possible to announce a definite outcome such that a suitable site will be allocated for the eventual creation of a worthy science centre. It is unlikely that this site will please everybody, but we have reason to believe that a large majority will welcome it in view of the many advantages secured. In any case we have the assurance that favourable consideration will be given by Ministers to transference to a better site, should such be found.

The course of the discussion showed that the co-operation of several scientific societies depended on that of the Royal Society, and your representatives accepted the flattering implications, though not without some hesitation. To speak directly: that means willingness, if necessary in the general interest, to leave Burlington House and to function as the heart of the Science Centre located elsewhere.

The elaboration of so far-reaching a scheme is not at all a simple matter and involves polypartite negotiations and agreements. The Fellows of the Society will be directly consulted on the major issue just so soon as it can be put to them in a precise form.

One possibly related matter may be adumbrated at this early stage, so that we may think about it.

The Report of Council mentions the reconstitution of a 'Rutherford Memorial Committee' which has various schemes under consideration. At the last meeting of this Committee I ventured to suggest that, as one aspect of the Memorial, a 'Rutherford Hall' of noble design should form a part of the Science Centre, and this was given general approval by the Committee and later by Council.

Unfortunately, I must now turn from these roseate dreams of future glory to an incident of the present which has rudely disturbed our peace of mind. It is possible that our information is incomplete, or even inaccurate, but it is probable that we have ascertained the more important facts of the case.

I refer to the report that eminent Russian biologists have been constrained to subscribe to interpretations of the data in the field of genetics which they had previously rejected, or perhaps had thought unworthy of serious consideration. According to *Pravda* 'The Academy of Sciences forgot that the most important Principle in Science is the Party Principle'. That is a forthright declaration which leaves little scope for ambiguity. The incident is evidently of political rather than of scientific importance, and the Royal Society is not concerned with politics.

We regret that the Academy of Sciences of Moscow has broken off its long correspondence with us. We trust that the new conditions will not seriously impede the advance of biological science for which such qualifications as 'Western' are as irrelevant as they would be for a multiplication table. We impute no blame and express no opinion as a body, but that does not mean that we must take no cognizance of the occurrence which may have some lessons for us, at least by way of analogy. For example, we may observe that Governments are not infallible, yet must be obeyed. This reflection should make us more than ever alert to preserve intact the prized freedom of science in our own domain. Actually no direct attack is likely here, and should the unexpected happen it will certainly not be along the lines of compelling us to espouse some particular scientific theory or doctrine. Conceivably it could take the more subtle form of control of the character and direction of our scientific work. There is immediate danger in the current deprecation of fundamental research, not of course absolutely, but relatively, in comparison with technological applications. I hope it will be the opinion of all Fellows that the Royal Society should take a leading part in upholding our ideals and in clarifying ideas on these topics and particularly in insisting on the vital role which the highest kind of disinterested investigation must play in the life of the community. It is certainly not sufficiently realized that the body scientific can only flourish when all its organs are in a healthy condition. As in a biological equilibrium there is a natural interdependence between pure and applied research. Pure science is fertilized by the advance of technology and vice versa. It would be quite consistent, though lamentable, to take up the position that we will have no more research at all and devote our energies to the exploitation of present knowledge. But it is impossible to dissect the elements of real progress. If we isolate one of the limbs of the organism it will not grow and will soon die.

I do not propose to add more than brief comment on new scientific discovery. It will be conceded that the great privilege of addressing the Fellows of the Royal Society is not without its own peculiar embarrassments. For example, I have recently delivered two lectures summarizing results of many years in my own field of work. If I venture on other ground I shall be speaking as an amateur to some of the leading specialists in the world.

Without attempting any kind of assessment of values, or any completeness of survey, I will mention a few things that have been brought to my notice in recent months, often as the result of casual conversation.

Two years ago I referred to the very great value of spectroscopic methods for the study of complex molecules of organic compounds. Since that time there have been striking applications of both ultra-violet and infra-red absorption measurements which have solved some otherwise intractable problems. But one difficulty that threatened to impede progress was the management of minute quantities of material, especially in the solid state.

Dr C. R. Burch, Warren Research Fellow, has built a reflecting microscope which is perfectly achromatic, and this has now been harnessed to an ultra-violet spectrometer by Barer, Holiday and Jope and to an infra-red spectrometer by Thompson, Barer and Cole. In the latter case the spectral range from 1 to 14μ can be covered, and the results are identical with those obtained with larger specimens in more usual apparatus.

The spectra of single biological cells or single crystals weighing less than 10^{-7} g. have been measured. Polarized infra-red radiation has also been used, and this promises to open up a new method for study of biological specimens, such as muscle fibre, in which orientation exists, as well as new possibilities for crystal analysis. This arises from the relations existing between absorption and orientation of the vibrating groups whence the angle between certain characteristic bonds and the crystal axis can be determined. Measurements have already been made with minute quantities of new antibiotics and similar substances in connexion with their identification, or the diagnosis of their molecular groups.

The importance of this advance for organic chemists and biochemists cannot be exaggerated. It will surely lead to the routine study of the infra-red absorption of many new compounds, and the relation between the spectra and constitution will therefore become still better known. As the method thus increases in power the demand for its help must also become more insistent.

I am most grateful to Dr H. W. Thompson for advance news of this remarkable development in technique, details of which will be published in a short time. Soon we shall all be repeating the wistful inquiry that ends Barrie's play: 'How much do they cost—those machines?'

I cannot pretend to be competent to discuss applied mathematics, astronomy, or physics, but it is obvious even to the uninitiated that great advances are being made in our knowledge of the mesons and of nuclear fission.

Sir George Thomson in his Bakerian Lecture gave a fascinating account of the evidence provided by study of the effects of cosmic rays on atoms, whilst Professor Lawrence with the aid of his new synchro-cyclotron has been able to produce mesons in the laboratory for the first time.

Professor E. A. Milne's forthcoming book will evoke lively discussion, for he tells me that acceptance of his theoretical system involves the assumption that Planck's constant h varies secularly with the time.

On the technical side a new type of valve which makes use of the semi-conducting properties of germanium has been introduced by Brattain & Bardeen of the Bell Telephone Co. Laboratories. This is regarded by the *cognoscenti* as an extremely significant development of electronic devices.

Coming nearer to my personal interests, the chemists and biologists of Parke Davis and Co. are to be warmly congratulated on their discovery and investigation of chloromycetin, a new antibiotic which has proved to be highly effective against scrub-typhus and certain analogous infections. The constitution of chloromycetin has not yet been disclosed, but it is said to be known, and apparently the substance has been synthesized. It makes possible the first satisfactory chemotherapy of a disease caused by a rickettsia and opens up the possibility of the control of virus diseases more generally.

Another discovery of vast medical, agricultural, and biochemical interest arises from the isolation of the intrinsic, anti-pernicious anaemia factor from liver. This brilliant consummation, a further triumph of modern chromatography, has been reached by A. Lester-Smith and his colleagues in the Glaxo Laboratories and by a team of chemists of Merck and Co. (New Jersey) led by K. Folkers. The red crystalline substance contains cobalt in a co-ordination complex which has a molecular weight of about 1600 and probably includes three atoms of phosphorus. It is proposed to call the factor vitamin B_{12}, and it is one of the most physiologically active of known substances, a dose of 1 μg. daily being fully effective therapeutically.

The picture is very quickly being filled in; we heard from Dr Marston quite recently of his outstanding researches on the cobalt requirements of sheep. A defect of cobalt in the soil leads to sickness and stunted growth with anaemia. Other recent and related work by Drs Tosic and Mitchell at Aberdeen concerns the assimilation of cobalt by micro-organisms. We may be sure that knowledge of the constitution of B_{12} will suggest other correlations, perhaps with nucleotides, perhaps with folic acid, which will be of vast significance. The molecule is not especially complex by modern standards, and its structure can be determined, provided sufficient effort is made. If I am reminded that we are unable to synthesize penicillin the reply would be that the problems are not comparable. Analysis of substances of this degree of complexity must nowadays always succeed, whereas synthesis may present a difficulty of a different order altogether. The trouble will be to get enough of the factor for investigation by ordinary methods, and this is a good illustration of the value of micro-physical devices. I am told that the infra-red absorption spectrum, determined by the new technique mentioned a few minutes ago, has already given valuable information, confirming the presence of PO, NH and OH groups, and indicating absence of aliphatic CH, but probable presence of aromatic CH as in benzene or purines.

We seem to be on the verge of great discoveries in another field perhaps not entirely unconnected with the last mentioned, that is, the problems connected with the effects of chemical substances on cell growth, including malignancy. There cannot be said to be any satisfactory treatment of cancer by chemical means, though various agents exert a beneficial effect in special cases. Such are stilboestrol, testosterone and the so-called nitrogen mustards. Of these, stilboestrol has the best claim to be regarded as a curative agent in a proportion of cases, and the halo-alkylamines ameliorate the condition of the patient. In the United States, Dr Rhoads and his colleagues have paid much attention to antagonists of folic acid. These are synthetic compounds, the molecules of which are planned to resemble those of folic acid. Several of them have been found to exert an inhibitory effect on the bone-marrow, producing anaemia, which can be reversed by folic acid. Similar antagonism is observed with bacteria, and, furthermore, Hertz has reported that certain of these substances inhibit the action of stilboestrol on the oviduct of the chick.

One of the most promising of the substances is called A-methopterin. This is folic acid with an amino-group replacing hydroxyl and a methyl group replacing hydrogen, quite a close analogue. The substance is anti-folic, anti-oestrogenic, and it has a definite inhibitory effect on the growth of tumours in laboratory animals, as well as of malignant cells in tissue culture. Folic acid is regarded as necessary for the growth of all cells, and the hope is to find a substance of low toxicity which will antagonize this factor at the level necessary to immobilize the more sensitive malignant growth without serious detriment to normal cells. It is known that laboratory work has encouraged clinical trials, but the results have not yet been disclosed.

Dr T. B. Heaton and my wife have published a preliminary account of work which finds its origin in observations made many years ago on a growth-inhibiting constituent of yeast (Heaton 1926). This was a differential growth inhibitor affecting the connective tissue cells but not the epithelia of cultures *in vitro*.

They have now prepared carbohydrate material from wheat middlings which is either itself active, or more probably contains an active principle. This produces retrogression of the implanted Walker carcinoma in rats.

It is decidedly interesting to find such a substance in a foodstuff; it may even be a vitamin. Since in some of the earlier experiments the material had been held at a low pH, the formation of hydroxymethylfurfuraldehyde was a possibility, and it was known that many aldehydes possessed growth-inhibitory properties. It was found that hydroxymethylfurfuraldehyde, or better its polymeride, also caused the retrogression of the tumours. The connexion, if any, between these findings is not yet clear.

At the Chester Beatty Research Institute Dr A. Haddow and his colleagues have tested the effects on normal growth of a number of pyrimidine derivatives. The developments have sprung from experiments initiated for other purposes—and the sequence is of more than passing interest. It started with the yellow enzyme, and a series of experiments designed to test the effect on growth of a series of synthetic analogues of the flavins.

Administration of 9-phenyl-5:6-benzo-*iso*-alloxazine to albino rats was accidentally, and most unexpectedly, found to produce an orange-yellow pigmentation of the hair (Haddow, Elson, Roe, Rudall & Timmis 1945). Pursuing this curious observation, other coloured substances of a similar type were tested, and among them xanthopterin, the butterfly-wing pigment, first isolated by Wieland & Schöpf in 1925. This did not colour the coat but produced a significant enlargement in the size of the rat kidney. It was an actual growth of the organ due to a great outburst of cell division in the kidney tubules.

In the course of still another research, that on the carcinogenicity of many styrylpyrimidines (Haddow, Kon & Ross), it was found that two substances of this group closely allied to xanthopterin in structure had the same effect in increasing the growth of the kidney. This cannot be mere coincidence, and proves that the effect is a primary one of xanthopterin and of the related pyrimidines on the cell. The phenomenon has been observed in the rat, mouse, *Peromyscus*, the rabbit, hamster and guinea-pig.

Now xanthopterin is of natural occurrence in the kidney, and it is possible that it is present as a growth regulator. Its structural relationship with folic acid will not be overlooked. Haddow and his associates feel that it is at least as important to study the means by which normal growth is so delicately adjusted as to inquire directly why the malignant cell is unregulated.

There are further experimental foundations for these ideas which lead Haddow and his collaborators to surmise that the co-ordinated growth of the normal tissues may depend upon the supply of essential substances from an external source, and that, contrariwise, malignant cells may have acquired the power to synthesize these essential substances, or their equivalents, themselves. Such suggestions are fully consonant with all that we know of the subject as a whole and give a new prospect for the chemistry of growth and differentiation.

I am much obliged to Dr Haddow for kindly acquainting me with results and ideas which are not yet published in full and for permission to mention them to-day.

I would not like to leave this subject without an incidental reference to Berenblum's significant discovery of sensitization of tissues to the action of carcinogenic agents. The constituents of croton oil, for example, directly, or indirectly, prepare the ground for a carcinogenic hydrocarbon in a remarkable manner. The underlying biochemical processes may be hard to bring to light, but the effort to do so would doubtless be repaid with interest.

As a drama like that of the pyrimidine group unfolds we can look back on the steps of the pioneers and be thankful that they selected the paths which led into such rich territory. But it often appeared to contemporaries that they were wandering into a barren wilderness. Whatever makes Hopkins think that there can be any possible interest in the pigments of butterflies? Why does Windaus waste his time and talent on that impossible substance cholesterol? I have heard those actual questions asked in past years, and would invite consideration of the answers that can be given to them now.

Another thought arising from recent progress is that the world of biochemistry, though of vast extent and interest, is finite, and we are beginning to sense the existence of its boundaries. We seem to encounter the same thing more often than might be expected; coincidences are indeed of frequent occurrence.

An example has already been mentioned, and another is the recent discovery that the photodynamic colouring matter of St John's wort, hypericin (Brockmann *et al.* 1939, 1942; Pace & McKinney 1941; Dhéré 1939, 1943) is closely allied to a pigment of the Aphididae, erythroaphin, studied by A. Todd and his collaborators at Cambridge (1948). And further, these substances are related to a mould pigment, oxypenicilliopsin, isolated by Oxford & Raistrick (1940). The pigments are *bis*-anthracene derivatives of some kind, and it is surprising to find these little-investigated and highly characteristic substances so widely distributed in nature.

A notable coincidence in my own field of work arose very recently from a theory of Woodward regarding the course of the synthesis of strychnine in the plant. He advanced what seemed at first the fantastic idea that a benzene nucleus, originally that of dihydroxyphenylalanine (or tyrosine), suffers fission so as to give two chains which enter into further transformations.

Transferring this mechanism piecemeal into an entirely different group of alkaloids, it was found to predict that constitution of emetine which can now be experimentally demonstrated to be correct. I am glad of this opportunity to say that I was unaware until quite recently of parallel, independent work by Dr H. T. Openshaw on the constitution of emetine. This was submitted for publication some months ago and includes the establishment of one detail of the structure not fully proved by the work of Späth, of Pailer and of Karrer. A few months ago no connexion whatever could be discerned between strychnine and emetine; they seemed poles apart. Now at one stroke they are connected by recognition of a common and remarkable type of biogenesis.

Flow of a compressible fluid about a cylinder
II. Flow with circulation

By T. M. Cherry, *The University of Melbourne*

(*Communicated by G. Temple, F.R.S.—Received* 6 *May* 1948)

This is a sequel to an earlier paper (Cherry 1947) in which was found a family of exact solutions for compressible flow past a cylinder. In the present paper the solution is extended to the case where the circulation round the cylinder is not zero. The formulae are developed for the case where the circulation is sufficiently small for the existence of a pair of stagnation points on the surface of the cylinder, under the condition that the speed at infinity is subsonic. One substantial point which arises in the present investigation is that the most direct generalization of the formulae for incompressible flow yield multiple-valued formulae for compressible flow. To get a single-valued solution it is necessary to add another multiple-valued solution, involving a set of constants which are to be determined from an infinite set of linear equations. The explicit solution of these equations is found, and hence the flow around a profile which is a slightly distorted circular cylinder.

Introduction

This is a sequel to a paper with the same title (Cherry 1947, referred to as part I), in which was found a family of exact solutions for compressible flow past a cylinder. In the present paper, the solution is extended to the case where the circulation round the cylinder is not zero.

The formulae are developed for the case where the circulation is sufficiently small for the existence of a pair of stagnation points on the surface of the cylinder. It is supposed that the speed at infinity q_1 is subsonic, but the speed elsewhere may be supersonic; indeed, the position co-ordinates and stream function are given in terms of the velocity (which is the independent variable) by functions which are everywhere regular, except for the singularities which a solution of the proposed character must necessarily have. But, of course, when q_1 is near the sonic value, the stream lines in the flow-plane may have cusps, as in the case of other solutions for mixed flows found by the hodograph method. The cylinder past which the fluid flows is approximately circular when q_1 is small; its shape for larger q_1 must be determined by numerical work which has not yet been undertaken.†

The work is in the main an application of the theory developed in part I, but one substantial new point arises. If we start from the formulae for incompressible flow with circulation and take their most direct generalization for compressible flow, we obtain a solution which is multiple-valued. To get a single-valued solution it is necessary to add another multiple-valued solution, involving a set of constants which are to be determined from an infinite set of linear equations; and the explicit solution of these equations is found.

---

† Concerning this shape, the only precise statements than can be made are that it has an axis of symmetry transverse to the stream, and that for small free-stream Mach numbers (for which there is a simpler approximate theory) the transverse axis is less than the longitudinal axis. Symmetry about a longitudinal axis is absent when the circulation is not zero. With this proviso, one may expect the shape to resemble what has been found by numerical work on the formulae of part I, of which a brief account (with figures) follows this paper on p. 32.

Section 1 gives a summary of the notation and of those results from part I which will often be required.

[*Added in proof*]. Lighthill (1947) has given formulae for circulatory flow past a cylinder, which have the advantage (over those here developed) of involving only singly-infinite series. His formulae are however singular at $\tau = 0$: in such manner that, if the generating function (i.e. the solution for a 'corresponding' incompressible flow) is regular, then in general the position co-ordinates x, y are logarithmically infinite for $\tau \sim 0$. (The critical point is that his formula for the stream function, when expressed as a series of elementary solutions, generally involves the term $\psi_1(\tau)\, e^{i\theta}$ or—in my notation—$\tau^{\frac{1}{2}} \tilde{F}_1(\tau)\, e^{-i\theta}$. So far as *one* point where $\tau \sim 0$ is concerned we can subtract the singular part of this solution and so obtain a regular stagnation point, but the singularity at any other point where $\tau \sim 0$ will in general remain. Hence the problem of the present paper, where there are two regular stagnation points, cannot be solved by his formulae.

1. SUMMARY OF PREVIOUS RESULTS†

Let the rectangular velocity components be $u = q\cos\theta$, $v = q\sin\theta$. For independent variables we take θ and $\tau = Cq^2$, where C is so chosen that $\tau = 1$ corresponds to the cavitation speed; to the sonic speed then corresponds a value τ_s which has the value 0·1684... when the adiabatic constant γ is 1·405. There is a potential Ω, for which the field equation is

$$4(1-\tau)\left(\tau^2\frac{\partial^2\Omega}{\partial\tau^2}+\tau\frac{\partial\Omega}{\partial\tau}\right)-4\beta\tau^2\frac{\partial\Omega}{\partial\tau}+(1-\tau-2\beta\tau)\frac{\partial^2\Omega}{\partial\theta^2}=0 \quad \left(\beta=\frac{1}{\gamma-1}\right), \tag{1·1}$$

from which the position co-ordinates x, y are found by differentiation:

$$\frac{x}{a}=\frac{\partial\Omega}{\partial(q\cos\theta)}, \quad \frac{y}{a}=\frac{\partial\Omega}{\partial(q\sin\theta)}, \tag{1·2}$$

or equivalently

$$\frac{x\sqrt{\tau}}{a'}=2\tau\frac{\partial\Omega}{\partial\tau}\cos\theta-\frac{\partial\Omega}{\partial\theta}\sin\theta, \quad \frac{y\sqrt{\tau}}{a'}=2\tau\frac{\partial\Omega}{\partial\tau}\sin\theta+\frac{\partial\Omega}{\partial\theta}\cos\theta, \tag{1·3}$$

where a, a' are scale constants.

The equation (1·1) is satisfied by

$$\Omega = \mathrm{e}^{\pm i\nu\theta}\tau^{\frac{1}{2}\nu}F(\tau), \tag{1·4}$$

provided $$\tau(1-\tau)F''+\{(1+\nu)(1-\tau)-\beta\tau\}F'+\tfrac{1}{2}\beta\nu(\nu-1)F=0. \tag{1·5}$$

If ν is not an integer the complete solution of (1·5) is

$$F = AF_\nu(\tau)+B\tau^{-\nu}F_{-\nu}(\tau),$$

where $F_\nu(\tau)$ is the hypergeometric function

$$F_\nu(\tau)=F(a_\nu, b_\nu, \nu+1, \tau)=1+\frac{a_\nu b_\nu}{\nu+1}\frac{\tau}{1!}+\frac{a_\nu(a_\nu+1)\,b_\nu(b_\nu+1)}{(\nu+1)(\nu+2)}\frac{\tau^2}{2!}+\dots, \tag{1·6}$$

---

† In one or two instances these results are only implicit in part I.

with parameters a_ν, b_ν given by

$$a_\nu + b_\nu = \nu + \beta, \quad a_\nu b_\nu = -\tfrac{1}{2}\beta\nu(\nu-1). \tag{1·7}$$

If ν is a positive integer, $F_\nu(\tau)$ remains an ordinary hypergeometric series, but $F_{-\nu}(\tau)$ is formally a series with zero denominators, and is to be replaced by a logarithmic solution $\mathscr{F}_{-\nu}(\tau)$ which will be found below.

For $\nu = 0$ we have $F_0(\tau) = 1$, and the solution (1·4) is useless. We may obtain useful solutions by putting in (1·1)

$$\Omega = \pm i\theta + \mathscr{G}_0(\tau); \tag{1·8}$$

it is easy to form and solve an equation for $\mathscr{G}_0(\tau)$, but we shall below obtain this function by another process.

The stream function ψ satisfies a field equation similar to (1·1), and there are elementary solutions

$$\psi = e^{\pm i\nu\theta}\tau^{\frac{1}{2}\nu}\tilde{F}_\nu(\tau), \tag{1·9}$$

where†

$$\tilde{F}_\nu(\tau) = F(\nu - a_\nu, \nu - b_\nu, \nu+1, \tau), \tag{1·10}$$

the F on the right being the hypergeometric function. But it is, of course, necessary that the solutions for ψ and Ω should correspond, so instead of appealing to the field equation for ψ we shall use the rule: *If*

$$\Omega = \sum_\nu (A_\nu e^{i\nu\theta} + \bar{A}_\nu e^{-i\nu\theta})\, \tau^{\frac{1}{2}\nu} F_\nu(\tau) \tag{1·11}$$

is a solution in which the range of values of ν *(generally infinite) does not include negative integers, then the most general corresponding* ψ *is given by*

$$i\psi = \sum_\nu (\nu - 1)(A_\nu e^{i\nu\theta} - \bar{A}_\nu e^{-i\nu\theta})\, \tau^{\frac{1}{2}\nu} \tilde{F}_\nu(\tau) + C, \tag{1·12}$$

where C *is an arbitrary constant.*

The analytic continuations of infinite series of the forms (1·11) and (1·12) are obtained via the partial fraction expansions (which we shall refer to as P.F.)

$$F_\nu(\tau) = T^\nu \left\{ \omega(\tau) - \sum_{m=1}^{\infty} \frac{h_m (\tau T)^m F_m(\tau)}{m+\nu} \right\}, \tag{1·13}$$

$$\tilde{F}_\nu(\tau) = T^\nu \left\{ \tilde{\omega}(\tau) - \sum_{m=1}^{\infty} \frac{\tilde{h}_m (\tau T)^m \tilde{F}_m(\tau)}{m+\nu} \right\}; \tag{1·14}$$

these are valid for $0 \leqslant \tau < \tau_s$, where $\tau_s = 1/(1+2\beta)$ corresponds to the sonic speed. Here

$$\left.\begin{aligned} \omega(\tau) = (1-\tau)^{\frac{1}{4}-\frac{1}{2}\beta}(1-\tau/\tau_s)^{-\frac{1}{4}}, \quad \tilde{\omega}(\tau) = (1-\tau)^{\frac{1}{4}+\frac{1}{2}\beta}(1-\tau/\tau_s)^{-\frac{1}{4}}, \\ T = T(\tau) = \frac{2\sqrt{\tau_s}}{\sqrt{\{\tau_s(1-\tau)\}} + \sqrt{(\tau_s - \tau)}} \left(\frac{\sqrt{(1-\tau)} + \sqrt{(\tau_s - \tau)}}{1 + \sqrt{\tau_s}} \right)^{1+2\alpha}, \end{aligned}\right\} \tag{1·15}$$

where α is the positive root of $\alpha^2 + \alpha = \frac{1}{2}\beta$; and the constants h_m, $\tilde{h}_m$ are given by

$$h_m = \frac{\Gamma(a_m)\,\Gamma(1+m-b_m)}{\Gamma(a_m - m)\,\Gamma(1-b_m)\,\Gamma(m)\,\Gamma(m+1)}, \quad \tilde{h}_m = \frac{m-1}{m+1} h_m. \tag{1·16}$$

---

† The notation $\overline{F}_\nu(\tau)$ of part I is here changed to $\widetilde{F}_\nu(\tau)$, in order that the bar notation may be used exclusively for complex conjugates.

For the purposes of (1·13) and (1·14) m is a positive integer, and this formula for h_m reduces to a form symmetrical in a_m, b_m; and, by (1·7), h_m is determined without ambiguity.

For discussion of the convergence of series such as (1·11), we require to know the order of magnitude of $F_\nu(\tau)$ when ν is large and real. If ν is excluded from ranges $(-n-\epsilon, -n+\epsilon)$ enclosing the negative integers $-n$, the results are:

$$|F_\nu(\tau)| \leqslant A\,|\nu|^{\frac{1}{2}}\,|T|^\nu, \quad A \text{ an abolute constant.} \tag{1·17}$$

$$\left.\begin{array}{l}\text{As } \tau \text{ increases from 0 to } \tau_s,\ \tau^{\frac{1}{2}}T \text{ increases from 0 to } \delta,\ = \alpha^\alpha/(1+\alpha)^{1+\alpha};\\ \text{as } \tau \text{ increases from } \tau_s \text{ to 1, } T \text{ becomes complex and } |\tau^{\frac{1}{2}}T| \text{ remains}\\ \text{equal to } \delta.\end{array}\right\} \tag{1·18}$$

For ν large positive we have

$$h_\nu = \frac{1}{2\pi\delta^{2\nu}}\left\{1+O\left(\frac{1}{\nu}\right)\right\}. \tag{1·19}$$

The generating function

Let τ_1 be a constant between 0 and τ_s, and let $T_1 = T(\tau_1)$ be the corresponding value of $T(\tau)$ as given by (1·15). Let

$$\zeta = \frac{\tau^{\frac{1}{2}}T\,\mathrm{e}^{i\theta}}{\tau_1^{\frac{1}{2}}T_1} = \frac{qT\,\mathrm{e}^{i\theta}}{q_1T_1}, \quad \bar{\zeta} = \frac{\tau^{\frac{1}{2}}T\,\mathrm{e}^{-i\theta}}{\tau_1^{\frac{1}{2}}T_1} = \frac{qT\,\mathrm{e}^{-i\theta}}{q_1T_1}. \tag{1·20}$$

As the standard forms of solution normalized with respect to τ_1 we take

$$\Omega = \sum_\nu C_\nu \mathrm{e}^{\nu i\theta}\left(\frac{\tau^{\frac{1}{2}}}{\tau_1^{\frac{1}{2}}T_1}\right)^\nu F_\nu(\tau) = \sum_\nu C_\nu\left(\frac{\zeta}{T}\right)^\nu F_\nu(\tau), \tag{1·21}$$

$$\Omega = \sum_\nu C_\nu \mathrm{e}^{-\nu i\theta}\left(\frac{\tau^{\frac{1}{2}}}{\tau_1^{\frac{1}{2}}T_1}\right)^\nu F_\nu(\tau) = \sum_\nu C_\nu\left(\frac{\bar{\zeta}}{T}\right)^\nu F_\nu(\tau), \tag{1·22}$$

where at present ν is excluded from negative integral values, and the C_ν are constants such as binomial coefficients. On substituting for the $F_\nu(\tau)$ from P.F. the solution (1·21) takes the form

$$\Omega = \omega(\tau)\,g(\zeta) - \sum_{m=1}^{\infty} h_m(\tau T)^m F_m(\tau)\,g_m(\zeta), \tag{1·23}$$

where

$$g(\zeta) = \sum_\nu C_\nu\zeta^\nu, \quad g_m(\zeta) = \sum_\nu C_\nu\zeta^\nu/(\nu+m). \tag{1·24}$$

We shall call $g(\zeta)$ the *generating function* of the solution (1·21); it can be interpreted as the solution for incompressible flow corresponding† to (1·21), with ζ as a normalized complex velocity variable. Similarly, (1·22) has the generating function $g(\bar{\zeta})$. We call $g_m(\zeta)$ the mth *associate* of $g(\zeta)$.

If we restrict τ, θ to be real, then to any range of τ in $(0, \tau_s)$ corresponds by (1·20) an annulus in the ζ-plane; and, conversely, to an annulus $\zeta' < |\zeta| < \zeta''$ for which $\zeta'' < \delta/(\tau_1^{\frac{1}{2}}T_1)$ corresponds a range $\tau' < \tau < \tau''$ in $(0, \tau_s)$. Then from (1·17), if the series

---

† This statement is intended to be suggestive; a precise statement would require a precise definition of 'corresponding' compressible and incompressible flows.

(1·24) for $g(\zeta)$ converges in the annulus, the series (1·21) converges absolutely for $\tau' < \tau < \tau''$; and the rearrangement whereby we pass from (1·21) to (1·23) is valid.

The continuation of the function defined by (1·21), when τ lies outside the range (τ', τ''), is obtained by finding series continuations for $g(\zeta)$, $g_m(\zeta)$ outside the annulus, substituting these in (1·23), and rearranging in the form (1·21).

2. LOGARITHMIC SOLUTIONS

We shall find solutions which answer to (1·4), when ν is a negative integer or zero. With a view to later application, it is convenient to work with solutions normalized with respect to a value τ_1 of τ, in the manner of (1·21). Let n be a positive integer, and let ν be non-integral. Then the field equation (1·1) has a solution

$$\Omega = (n+\nu)\,\mathrm{e}^{i\nu\theta}\left(\frac{\tau^{\frac{1}{2}}}{\tau_1^{\frac{1}{2}}T_1}\right)^{\nu} F_\nu(\tau) = (n+\nu)\left(\frac{\zeta}{T}\right)^{\nu} F_\nu(\tau); \tag{2·1}$$

from the series (1·6) for $F_\nu(\tau)$ in powers of τ, this solution is a function of ν regular at $\nu = -n$; in particular

$$\lim_{\nu\to -n}\left(\frac{\zeta}{T}\right)^{\nu}(n+\nu)\,F_\nu(\tau) = -h_n\left(\frac{\tau T}{\zeta}\right)^{n} F_n(\tau) = -h_n\,T_1^n(\tau\tau_1)^{\frac{1}{2}n}\,\mathrm{e}^{-in\theta}\,F_n(\tau), \tag{2·2}$$

as may be verified immediately from (1·13). Since the field equation does not involve ν, we may derive the new solution

$$\begin{aligned}\Omega = L_n(\zeta) &= \lim_{\nu\to -n}\frac{\partial}{\partial\nu}\left[(n+\nu)\left(\frac{\zeta}{T}\right)^{\nu} F_\nu(\tau)\right]\\ &= \log\frac{\zeta}{T}\lim\,(n+\nu)\left(\frac{\zeta}{T}\right)^{\nu} F_\nu(\tau) + \lim\left(\frac{\zeta}{T}\right)^{\nu}\frac{\partial}{\partial\nu}[(n+\nu)\,F_\nu(\tau)],\end{aligned}$$

i.e.

$$L_n(\zeta) = -\log\frac{\zeta}{T}\,h_n\left(\frac{\tau T}{\zeta}\right)^{n} F_n(\tau) + \left(\frac{\zeta}{T}\right)^{-n}\left[\frac{\partial}{\partial\nu}(n+\nu)\,F_\nu(\tau)\right]_{\nu=-n}, \tag{2·3}$$

by (2·2). L_n is a function of the variables τ, θ and of the parameter τ_1, but we may call it $L_n(\zeta)$ without ambiguity provided we refer to the form (1·20) for ζ and not merely to the value of ζ.

For $0 \leqslant \tau < \tau_s$ we have, on substituting for $F_\nu(\tau)$ from P.F.,

$$\begin{aligned}L_n(\zeta) &= \lim_{\nu\to -n}\frac{\partial}{\partial\nu}\left[(n+\nu)\,\zeta^{\nu}\left\{\omega(\tau) - \sum_m \frac{h_m(\tau T)^m F_m(\tau)}{m+\nu}\right\}\right]\\ &= \log\zeta\,\lim\,(n+\nu)\,\zeta^{\nu}\left\{\omega(\tau) - \sum_m \frac{h_m(\tau T)^m F_m(\tau)}{m+\nu}\right\} + \lim\,\zeta^{\nu}\left\{\omega(\tau) - \sum_m{}' \frac{h_m(\tau T)^m F_m(\tau)}{m+\nu}\right\},\end{aligned}$$

i.e. (using (2·2)),

$$L_n(\zeta) = -\log\zeta\,h_n\left(\frac{\tau T}{\zeta}\right)^{n} F_n(\tau) + \zeta^{-n}\left\{\omega(\tau) - \sum_{m=1}^{\infty}{}' \frac{h_m(\tau T)^m F_m(\tau)}{m-n}\right\}, \tag{2·4}$$

where $\sum_m{}'$ omits the term with $m = n$.

Another solution for Ω is given by

$$\Omega = M_n(\zeta) = \lim_{\nu\to n}\frac{\partial}{\partial\nu}\left[\left(\frac{\zeta}{T}\right)^{\nu} F_\nu(\tau)\right] = \log\frac{\zeta}{T}\left(\frac{\zeta}{T}\right)^{n} F_n(\tau) + \left(\frac{\zeta}{T}\right)^{n}\left[\frac{\partial F_\nu(\tau)}{\partial\nu}\right]_{\nu=n}. \tag{2·5}$$

For $0 \leqslant \tau < \tau_s$, the partial fraction expansion is, from P.F.,

$$\begin{aligned} M_n(\zeta) &= \lim_{\nu \to n} \frac{\partial}{\partial \nu} \zeta^\nu \left\{ \omega(\tau) - \sum_m \frac{h_m (\tau T)^m F_m(\tau)}{m+\nu} \right\} \\ &= \log \zeta \left(\frac{\zeta}{T} \right)^n F_n(\tau) + \zeta^n \sum_{m=1}^{\infty} \frac{h_m (\tau T)^m F_m(\tau)}{(m+n)^2}. \end{aligned} \tag{2·6}$$

Since $\log \zeta = i\theta + \log (\tau^{\frac{1}{2}} T / \tau_1^{\frac{1}{2}} T_1)$, the solutions $L_n(\zeta)$, $M_n(\zeta)$ are not periodic in θ. To obtain a periodic solution we note that

$$\zeta \bar{\zeta} = \tau T^2 / \tau_1 T_1^2, \tag{2·7}$$

so that $\tau T / \zeta = \tau_1 T_1^2 \bar{\zeta} / T$. Hence combining $L_n(\zeta)$ with the conjugate of $M_n(\zeta)$, we have from (2·3) and (2·5) the solution

$$\begin{aligned} \Omega &= L_n(\zeta) - h_n (\tau_1 T_1^2)^n M_n(\bar{\zeta}) \\ &= -h_n \left(\frac{\tau T}{\zeta} \right)^n F_n(\tau) \log \frac{\zeta \bar{\zeta}}{T^2} + \left(\frac{\zeta}{T} \right)^{-n} \left[\frac{\partial}{\partial \nu} (n+\nu) F_\nu(\tau) \right]_{\nu=-n} - h_n \left(\frac{\tau_1 T_1^2 \bar{\zeta}}{T} \right)^n \left[\frac{\partial F_\nu(\tau)}{\partial \nu} \right]_{\nu=n} \\ &= \left(\frac{\zeta}{T} \right)^{-n} \left\{ h_n \tau^n F_n(\tau) \log (\tau_1 T_1^2) - h_n F_n(\tau) \left[\frac{\partial \tau^\nu}{\partial \nu} \right]_{\nu=n} \right. \\ &\qquad \left. - h_n \tau^n \left[\frac{\partial F_\nu(\tau)}{\partial \nu} \right]_{\nu=n} + \left[\frac{\partial}{\partial \nu} (n+\nu) F_\nu(\tau) \right]_{\nu=-n} \right\} \\ &= \left(\frac{\zeta}{T} \right)^{-n} \left\{ h_n \tau^n F_n(\tau) \log (\tau_1 T_1^2) + \left[\frac{\partial}{\partial \nu} \{ (n+\nu) F_\nu(\tau) + h_n \tau^{-\nu} F_{-\nu}(\tau) \} \right]_{\nu=-n} \right\} \\ &= \left(\frac{\zeta}{T} \right)^{-n} \left\{ h_n \tau^n F_n(\tau) \log (\tau_1 T_1^2) + \lim_{\nu \to -n} \left[F_\nu(\tau) + \frac{h_n \tau^{-\nu} F_{-\nu}(\tau)}{n+\nu} \right] \right\}, \end{aligned}$$

the last line being a consequence of (2·2).

Since $(\zeta/T)^{-n} = (\tau_1^{\frac{1}{2}} T_1)^n \tau^{-\frac{1}{2}n} \mathrm{e}^{-ni\theta}$, we see that the function in $\{\ldots\}$ must satisfy the hypergeometric equation (1·5) with $\nu = -n$. Hence

$$\mathscr{F}_{-n}(\tau) = \lim_{\nu \to -n} \left[F_\nu(\tau) + \frac{h_n \tau^{-\nu} F_{-\nu}(\tau)}{n+\nu} \right] \tag{2·8}$$

is a logarithmic solution which answers to $F_\nu(\tau)$ when $\upsilon = -n$; and the result just obtained is

$$L_n(\zeta) - h_n (\tau_1 T_1^2)^n M_n(\bar{\zeta}) = \left(\frac{\zeta}{T} \right)^{-n} [\mathscr{F}_{-n}(\tau) + h_n \tau^n F_n(\tau) \log (\tau_1 T_1^2).] \tag{2·9}$$

From (2·8) we can derive, by routine methods, the series[†]

$$\begin{aligned} \mathscr{F}_{-n}(\tau) = 1 &+ \frac{(a_n - n)(n - b_n)}{n-1} \frac{\tau}{1!} + \ldots + \frac{(a_n - n) \ldots (a_n - 1)(n - b_n) \ldots (1 - b_n)}{(n-1) \ldots 1} \frac{\tau^{n-1}}{(n-1)!} \\ &+ \sum_{p=1}^{\infty} \frac{\Gamma(a_n + p) \Gamma(n + 1 - b_n)}{\Gamma(a_n - n) \Gamma(1 - b_n - p)} \frac{(-1)^p \tau^{n+p}}{(n-1)! (n+p)! p!} \\ &\times \{ \psi(1 - b_n) - \psi(1 - b_n - p) + \psi(a_n) - \psi(a_n + p) + \psi(p+1) - \psi(1) \\ &+ \psi(n+p+1) - \psi(n+1) \} + h_n \eta_n \tau^n F_n(\tau) - h_n \tau^n F_n(\tau) \log \tau, \end{aligned} \tag{2·10}$$

[†] When $n > 1$, the relations (1·7) give a_n, b_n real and of opposite signs. Choosing $a_n > 0$ and $b_n < 0$, the series (2·10) has been so written that the arguments of the Γ and ψ functions may be as far as possible positive; the only argument that can be non-positive is $1 - b_n - p$ and when this is zero or a negative integer (which arises when b_n is a negative integer and p is a sufficiently large integer) the infinity of $\psi(1 - b_n - p)$ is cancelled by the zero of $1/\Gamma(1 - b_n - p)$.

where $$\eta_n = \psi(n+1)+\psi(1)-\psi(a_n)-\psi(1-b_n)+d\log h_n/dn,$$

and $$\psi(z) = \Gamma'(z)/\Gamma(z).$$

Also from (2·9), with (2·4) and (2·6), we have for $0 \leqslant \tau < \tau_s$ the partial fraction expansion

$$\mathscr{F}_{-n}(\tau) = T^{-n}\left\{\omega(\tau) - \sum_{m=1}^{\infty}{}' \frac{h_m(\tau T)^m F_m(\tau)}{m-n}\right\}$$
$$-h_n \tau^n F_n(\tau) \log(\tau T^2) - h_n \sum_{m=1}^{\infty} \frac{h_m(\tau T)^{m+n} F_m(\tau)}{(m+n)^2}. \qquad (2{\cdot}11)$$

The order of magnitude of $\mathscr{F}_{-n}(\tau)$, when n is large, is given by

$$|\mathscr{F}_{-n}(\tau)| \leqslant A n^{\frac{1}{2}} |T|^{-n}, \qquad (2{\cdot}12)$$

a result which answers exactly to (1·17). The proof of this is deferred to another paper.

The function $M_n(\zeta)$ gives for $n = 0$

$$M_0(\zeta) = \log\frac{\zeta}{T} + \left[\frac{\partial F_\nu(\tau)}{\partial \nu}\right]_{\nu=0} = \log\zeta + \sum_{m=1}^{\infty} \frac{h_m(\tau T)^m F_m(\tau)}{m^2}, \qquad (2{\cdot}13)$$

and $\Omega = M_0(\zeta)$, $\Omega = M_0(\bar{\zeta})$ are two independent solutions to take the place of (1·4) for $\nu = 0$. Since $\log(\zeta/T) = i\theta + \frac{1}{2}\log\tau - \frac{1}{2}\log(\tau_1 T_1^2)$, these solutions are of the form (1·8), and suppressing the normalization constant $\frac{1}{2}\log(\tau_1 T_1^2)$, we can put

$$\mathscr{G}_0(\tau) = \tfrac{1}{2}\log\tau + \left[\frac{\partial F_\nu(\tau)}{\partial \nu}\right]_{\nu=0} = \tfrac{1}{2}\log\tau + \frac{1}{2}\int_0^\tau \{(1-\tau)^{-\beta} - 1\}\, d\tau/\tau.$$

By similar processes we may obtain logarithmic solutions $\tilde{L}_n(\zeta)$, etc., for ψ; the formulae are obtained from the preceding if we replace F_ν, h_ν, ω by $\tilde{F}_\nu$, $\tilde{h}_\nu$, $\tilde{\omega}$, so that, for instance, corresponding to (2·4),

$$\tilde{L}_n(\zeta) = -\log\zeta\, \tilde{h}_n \left(\frac{\tau T}{\zeta}\right)^n \tilde{F}_n(\tau) + \zeta^{-n}\left\{\tilde{\omega}(\tau) - \sum_{m=1}^{\infty}{}' \frac{\tilde{h}_m(\tau T)^m \tilde{F}_m(\tau)}{m-n}\right\}, \qquad (2{\cdot}14)$$

and corresponding to (2·8)

$$\tilde{\mathscr{F}}_{-n}(\tau) = \lim_{\nu \to -n}\left\{\tilde{F}_\nu(\tau) + \frac{\tilde{h}_n \tau^{-\nu} \tilde{F}_{-\nu}(\tau)}{\nu+n}\right\}. \qquad (2{\cdot}15)$$

Since $\tilde{h}_1 = 0$, we have

$$\tilde{\mathscr{F}}_{-1}(\tau) = \lim_{\nu \to -1} \tilde{F}_\nu(\tau) = 1 + \tfrac{1}{2}\beta\int_0^\tau (1-\tau)^\beta\, d\tau;$$

but this simplification is of no importance in the present application.

The rule whereby a series (1·12) for $i\psi$ is derived from a series (1·11) for Ω may now be extended to the case where the Ω series contains terms of the form

$$\Omega_{-n} = (A_{-n}\, \mathrm{e}^{-ni\theta} + \bar{A}_{-n}\, \mathrm{e}^{ni\theta})\, \tau^{-\frac{1}{2}n} \mathscr{F}_{-n}(\tau). \qquad (2{\cdot}16)$$

By appealing to the definitions (2·8) and (2·15) of $\mathscr{F}_{-n}(\tau)$, $\tilde{\mathscr{F}}_{-n}(\tau)$ it is found that (2·16) leads to $\psi = \psi_{-n}$, where

$$i\psi_{-n} = (\bar{A}_{-n}\, \mathrm{e}^{ni\theta} - A_{-n}\, \mathrm{e}^{-ni\theta})\left\{(n+1)\, \tau^{-\frac{1}{2}n} \tilde{\mathscr{F}}_{-n}(\tau) - \frac{2h_n}{n+1} \tau^{\frac{1}{2}n} \tilde{F}_n(\tau)\right\} + C. \qquad (2{\cdot}17)$$

We shall not need to use this rule, and omit the proof.

To conclude this section, the obvious remark is made that, as a fundamental pair of solutions of (1·5) for the case where ν is a negative integer $-n$, we can take $\tau^n F_n(\tau)$ along with any combination of the form $\mathscr{F}_{-n}(\tau)+A\tau^n F_n(\tau)$. In the present application it is most convenient to take $A=0$, but for other applications other choices of A may be better. For example, it may be proved that

$$G_\nu(\tau) = F_{-\nu}(\tau)+\frac{\Gamma(-\nu)\,\Gamma(a_\nu)\,\Gamma(b_\nu)}{\Gamma(\nu)\,\Gamma(a_\nu-\nu)\,\Gamma(b_\nu-\nu)}\,\tau^\nu F_\nu(\tau)$$

gives $$G_n(\tau) = \lim_{\nu\to n} G_\nu(\tau) = \mathscr{F}_{-n}(\tau)+(h_n\pi\cot b_n\pi-h'_n)\,\tau^n F_n(\tau).$$

The advantage of $G_\nu(\tau)$ is that it is a simple function of $1-\tau$,

$$G_\nu(\tau) = \frac{\Gamma(a_\nu)\,\Gamma(b_\nu)}{\Gamma(\beta)\,\Gamma(\nu)}\,F(a_\nu-\nu,b_\nu-\nu,\beta,1-\tau),$$

but it has the disadvantage of becoming infinite when b_n is a negative integer.

3. COMPRESSIBLE FLOW WITH CIRCULATION: TRIAL SOLUTION

At the end of § 1 it was stated that any solution for compressible flow is associated with a generating function, which is the solution for a 'corresponding' case of incompressible flow. Using the correspondence in the inverse direction, we shall find the potential Ω for incompressible flow, with circulation, past a circular cylinder, and we shall then consider the solution for compressible flow having this as generating function.

For the incompressible flow, the complex velocity potential is

$$\phi+i\psi = Q\left(z+\frac{a^2}{z}-2ika\log\frac{z}{a}\right),$$

where $z=x+iy$, a is the radius of the cylinder, k is a circulation constant, and Q is the velocity at infinity (which is parallel to Ox). Hence

$$\bar\zeta = \frac{q\,e^{-i\theta}}{Q} = \frac{1}{Q}\frac{d}{dz}(\phi+i\psi) = 1-\frac{a^2}{z^2}-\frac{2ika}{z}. \tag{3·1}$$

Solving for z/a we get

$$\frac{z}{a} = i\,\frac{k+\sqrt{(\bar\zeta+k^2-1)}}{1-\bar\zeta} = \frac{}{k-\sqrt{(\bar\zeta+k^2-1)}}, \tag{3·2}$$

where the radical is double-valued.

The stagnation points $\bar\zeta=0$ are at $z/a=ik\pm\sqrt{(1-k^2)}$. They lie on the surface of the cylinder $|z|=a$ if $0<k<1$, and we shall confine attention to this case. We write

$$k'=\sqrt{(1-k^2)},\quad 0<k'<1. \tag{3·3}$$

The potential Ω is determined by (1·2). Since here we have $q\,e^{-i\theta}=Q\bar\zeta$, $q\,e^{i\theta}=Q\zeta$, equivalent conditions for Ω are

$$\frac{\partial\Omega}{\partial\zeta}=\tfrac{1}{2}Q\bar z,\quad \frac{\partial\Omega}{\partial\bar\zeta}=\tfrac{1}{2}Qz,$$

where the relation between $\bar{z}$ and ζ is the conjugate of (3·2), viz.

$$\frac{\bar{z}}{a} = -i\frac{k+\sqrt{(\zeta-k'^2)}}{1-\zeta}. \tag{3·4}$$

Hence we find

$$\Omega = iQa[\sqrt{(\zeta-k'^2)}+k\log\{k-\sqrt{(\zeta-k'^2)}\}]-iQa[\sqrt{(\bar{\zeta}-k'^2)}+k\log\{k-\sqrt{(\bar{\zeta}-k'^2)}\}].$$

It is convenient to suppress the constant $-\frac{1}{2}iQa$, and we shall accordingly consider the compressible solution, i.e. the solution of (1·1), whose generating function is the expansion, in a suitably chosen domain, of

$$g(\zeta) = -2\sqrt{(\zeta-k'^2)}-2k\log\{k-\sqrt{(\zeta-k'^2)}\}. \tag{3·5}$$

A physical solution will be obtained if, at the end, we multiply by an arbitrary pure-imaginary factor and then add the conjugate solution.

The function $g(\zeta)$ just defined has various expansions, according to the range of ζ and the determination of the radical and logarithm. We shall first specify the ζ-domains in which the various expansions are valid, and shall indicate the corresponding regions in the plane of the incompressible flow; since the formulae for compressible flow merge continuously into those for incompressible flow as the speed tends to zero, this correspondence between ζ-domains and the flow-plane will be qualitatively valid in the 'slow' compressible case.

The relation (3·1) or (3·2) defines a one-one correspondence between the z-plane and a two-sheeted Riemann surface for $\bar{\zeta}$, having branch points at $\bar{\zeta} = 1-k^2 = k'^2$ and $\bar{\zeta} = \infty$; since ζ is the conjugate of $\bar{\zeta}$, the correspondence between a two-sheeted surface for ζ and the z-plane then follows immediately.

The $\bar{\zeta}$-surface is separated by the locus $Im\bar{\zeta} = 0$ into four half-planes. From (3·1) it follows easily that to the locus $Im\bar{\zeta} = 0$ corresponds in the z-plane the line $x = 0$ and the circle $x^2+y^2 = ay/k$, and the four regions into which the z-plane is thereby divided correspond to the four $\bar{\zeta}$-half-planes. This correspondence is indicated in the figure, where for definiteness we may suppose the pair-wise joining of the $\bar{\zeta}$-half-planes to be direct across the segments $(-\infty, k'^2)$ and crosswise across the segments $(k'^2, +\infty)$. Figure 1 shows also the loci $|\bar{\zeta}| = k'^2$, $|\bar{\zeta}| = 1$, which divide the $\bar{\zeta}$-surface into four domains; two of these are circular disks II, III, one in each sheet, the third is a two-sheeted ring I + I′, and the fourth is a two-sheeted ring IV + IV′; as the dividing line between I, I′ and between IV, IV′ we take the segment $(-\infty, -k'^2)$ of the locus $Im\bar{\zeta} = 0$. In figure 1, corresponding points of the $\bar{\zeta}$-surface and z-plane are similarly lettered.

We see now that in II we are to take

$$\sqrt{(\bar{\zeta}-k'^2)} = -ik'(1-\bar{\zeta}/k'^2)^{\frac{1}{2}}, \quad \sqrt{(\zeta-k'^2)} = ik'(1-\zeta/k'^2)^{\frac{1}{2}}, \tag{3·6}$$

where $(\ldots)^{\frac{1}{2}}$ denotes the principal value; for this gives in (3·2) and (3·4)

$$\frac{z}{a} = ik+k', \quad \frac{\bar{z}}{a} = -ik+k' \quad \text{for} \quad \zeta = \bar{\zeta} = 0,$$

i.e. the stagnation point S_2. Similarly in III we are to take

$$\sqrt{(\bar{\zeta}-k'^2)} = ik'(1-\bar{\zeta}/k'^2)^{\frac{1}{2}}, \quad \sqrt{(\zeta-k'^2)} = -ik'(1-\zeta/k'^2)^{\frac{1}{2}}. \tag{3·7}$$

In I and IV we are to take

$$\sqrt{(\bar\zeta-k'^2)}=\bar\zeta^{\frac12}(1-k'^2/\bar\zeta)^{\frac12},\quad \sqrt{(\zeta-k'^2)}=\zeta^{\frac12}(1-k'^2/\zeta)^{\frac12},\tag{3·8}$$

where $\bar\zeta^{\frac12}$, etc., denote principal values; for (3·2) and (3·4) then give, for ζ, $\bar\zeta$ near 1,

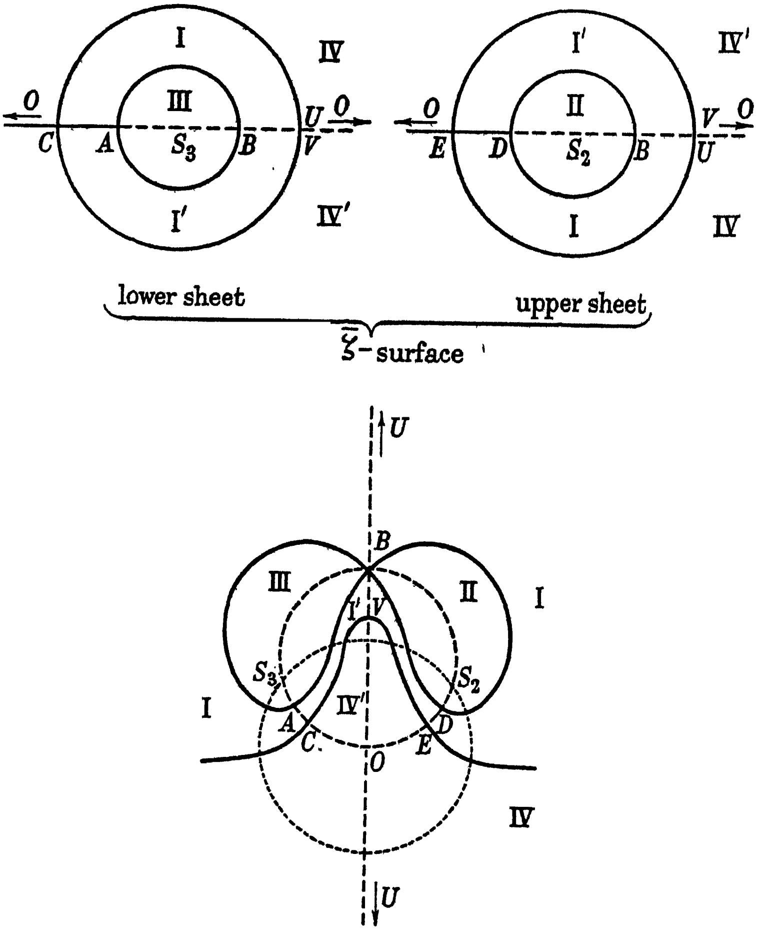
$$\frac{z}{a}=\frac{2ik}{1-\bar\zeta}+O(1),\quad \frac{\bar z}{a}=-\frac{2ik}{1-\zeta}+O(1),$$

so that the point U in I+IV corresponds to $z=\infty$. Similarly in I′ and IV′ we are to take

$$\sqrt{(\bar\zeta-k'^2)}=-\bar\zeta^{\frac12}(1-k'^2/\bar\zeta)^{\frac12},\quad \sqrt{(\zeta-k'^2)}=-\zeta^{\frac12}(1-k'^2/\zeta)^{\frac12};$$

for $\bar\zeta$ near 1 we then have

$$a \quad 2k+O(\bar\zeta-1)' \quad a \quad 2k+O(\zeta-1)'$$



z-plane

FIGURE 1. The curve on the $\bar\zeta$-surface corresponding to the circle $|z|=a$ (dotted thus: ......) is not shown.

so that the point V in $\mathrm{I}'+\mathrm{IV}'$ corresponds to $z = ia/2k$, a point which is above the origin, and is outside the circle $|z| = a$ if $k < \frac{1}{2}$.

The function $g(\zeta)$ defined by (3·5) is now determinate in the six domains I, ..., IV′, except for the determination of the logarithm; for this we choose the principal value on the segment $k'^2 < \zeta < 1$ of I, where its argument $k - \sqrt{(\zeta - k'^2)}$ is real positive, and its determination elsewhere follows by continuity for displacement of ζ along any assigned path. The corresponding expansions of $g(\zeta)$ are now determinate, as follows; in order not to interrupt the main argument, their proof is given separately in § 8:

$$\left.\begin{aligned}
g(\zeta) &= k\sum_{n=1}^{\infty}\frac{\zeta^n}{n} + k\sum_{n=1}^{\infty}\frac{\zeta^{n+\frac{1}{2}}}{n+\frac{1}{2}} + \sum_{n=0}^{\infty}\frac{q_n}{\frac{1}{2}-n}\zeta^{\frac{1}{2}-n}, && \text{(I)}\\
g(\zeta) &= k\sum_{1}^{\infty}\frac{\zeta^n}{n} - k\sum_{1}^{\infty}\frac{\zeta^{n+\frac{1}{2}}}{n+\frac{1}{2}} - \sum_{0}^{\infty}\frac{q_n}{\frac{1}{2}-n}\zeta^{\frac{1}{2}-n}, && \text{(I}'\text{)}\\
g(\zeta) &= -iC + \sum_{1}^{\infty}\frac{k+ik'r_n}{n}\zeta^n, && \text{(II)}\\
g(\zeta) &= iC + \sum_{1}^{\infty}\frac{k-ik'r_n}{n}\zeta^n, && \text{(III)}\\
g(\zeta) &= -k(2\pi i + \log\zeta) + k\sum_{1}^{\infty}\frac{\zeta^{-n}}{n} - \sum_{0}^{\infty}\frac{s_n}{\frac{1}{2}-n}\zeta^{\frac{1}{2}-n}, && \text{(IV)}\\
g(\zeta) &= -k\log\zeta + k\sum_{1}^{\infty}\frac{\zeta^{-n}}{n} + \sum_{0}^{\infty}\frac{s_n}{\frac{1}{2}-n}\zeta^{\frac{1}{2}-n}, && \text{(IV}'\text{)}
\end{aligned}\right\} \tag{3·9}$$

Here

$$C = 2(k' - k\arcsin k'), \quad q_n = \sum_{m=n+1}^{\infty} c_m k'^{2m}, \quad r_n = \sum_{m=0}^{n-1} c_m k'^{-2m}, \quad s_n = \sum_{m=0}^{n} c_m k'^{2m}, \tag{3·10}$$

where the c_m are binomial coefficients defined by

$$\sqrt{(1-\zeta)} = \sum_{m=0}^{\infty} c_m \zeta^m. \tag{3·11}$$

The expansions given for I′, II, III are valid when ζ moves from I into these domains along any path on which $|\zeta| < 1$; for the expansion IV we suppose that ζ has moved into the domain IV from I by circling round $\zeta = 1$ in the positive sense, so that on the common boundary of the domains, $\arg(\zeta-1)_{\mathrm{IV}} = \arg(1-\zeta)_{\mathrm{I}} - \pi$; for the expansions IV′ we suppose that ζ has moved into the domain IV′ by crossing the circle $|\zeta| = 1$ from the domain I′.

We now make what is logically a fresh start, by proposing for consideration the following solution of the equation (1·1) for compressible flow:

$$\begin{aligned}
\Omega = \Omega_0 = k\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{\tau^{\frac{1}{2}}\,\mathrm{e}^{i\theta}}{\tau_1^{\frac{1}{2}}T_1}\right)^n F_n(\tau) + k\sum_{n=1}^{\infty}\frac{1}{n+\frac{1}{2}}\left(\frac{\tau^{\frac{1}{2}}\,\mathrm{e}^{i\theta}}{\tau_1^{\frac{1}{2}}T_1}\right)^{n+\frac{1}{2}} F_{n+\frac{1}{2}}(\tau)&\\
+ \sum_{n=0}^{\infty}\frac{q_n}{\frac{1}{2}-n}\left(\frac{\tau^{\frac{1}{2}}\,\mathrm{e}^{i\theta}}{\tau_1^{\frac{1}{2}}T_1}\right)^{\frac{1}{2}-n} F_{\frac{1}{2}-n}(\tau),&
\end{aligned} \tag{3·12}$$

where $0 < \tau_1 < \tau_s$. Writing

$$\zeta = \frac{\tau^{\frac{1}{2}}T}{\tau_1^{\frac{1}{2}}T_1}\mathrm{e}^{i\theta}, \tag{3·13}$$

and substituting for the hypergeometric functions from P.F., this solution can be written

$$\Omega_0 = \omega(\tau)\, g(\zeta) - \sum_{m=1}^{\infty} h_m (\tau T)^m F_m(\tau)\, g_m(\zeta), \tag{3·14}$$

where the generating function $g(\zeta)$ is defined by (3·9, I), and $g_m(\zeta)$ is its mth associate.

The hypothesis $0 < \tau_1 < \tau_s$ secures that (3·12) converges absolutely when θ is real and τ lies in the range (τ', τ_1), which corresponds by (3·13) to the range $(k'^2, 1)$ of $|\zeta|$; and the rearrangement of double series leading to (3·14) is then valid.

It will save later repetition to consider Ω_0 as the case $p = 0$ of the solution

$$\Omega_p = k \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{\zeta}{T}\right)^{n+p} F_{n+p}(\tau) + k \sum_{n=1}^{\infty} \frac{1}{\frac{1}{2}+n} \left(\frac{\zeta}{T}\right)^{\frac{1}{2}+n+p} F_{\frac{1}{2}+n+p}(\tau)$$
$$+ \sum_{n=0}^{\infty} \frac{q_n}{\frac{1}{2}-n} \left(\frac{\zeta}{T}\right)^{\frac{1}{2}-n+p} F_{\frac{1}{2}-n+p}(\tau), \tag{3·15}$$

where p is a non-negative integer. We proceed to consider the analytic continuation of Ω_p, when ζ circles once in the positive sense round $\zeta = 1$, from I through IV and back to the starting point in I. From P.F. we have

$$\Omega_p = \omega(\tau)\, g^{(p)}(\zeta) - \sum_{m=1}^{\infty} h_m (\tau T)^m F_m(\tau)\, g_m^{(p)}(\zeta), \tag{3·16}$$

where

$$\begin{aligned} g^{(p)}(\zeta) &= k \sum_{1}^{\infty} \frac{\zeta^{n+p}}{n} + k \sum_{1}^{\infty} \frac{\zeta^{\frac{1}{2}+n+p}}{\frac{1}{2}+n} + \sum_{0}^{\infty} \frac{q_n}{\frac{1}{2}-n} \zeta^{\frac{1}{2}-n+p} \\ &= \zeta^p g(\zeta), \end{aligned} \tag{3·17}$$

$$\begin{aligned} g_m^{(p)}(\zeta) &= k \sum_{1}^{\infty} \frac{\zeta^{n+p}}{n(n+p+m)} + k \sum_{1}^{\infty} \frac{\zeta^{\frac{1}{2}+n+p}}{(\frac{1}{2}+n)(\frac{1}{2}+n+p+m)} + \sum_{0}^{\infty} \frac{q_n \zeta^{\frac{1}{2}-n+p}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)} \\ &= \zeta^{-m} \left\{ \int_{\zeta_0}^{\zeta} z^{p+m-1} g(z)\, dz + C_0 \right\}, \end{aligned} \tag{3·18}$$

ζ_0 being a point of I and C_0 a constant depending on ζ_0.

The continuation of Ω_p is formed by substituting in (3·16) the continuations of $g^{(p)}(\zeta)$, $g_m^{(p)}(\zeta)$. In I, near $\zeta = 1$, we have, expanding in powers of $\zeta - 1$,

$$\sqrt{(\zeta - k'^2)} = \sqrt{(\zeta + k^2 - 1)} = k + \tfrac{1}{2}(\zeta - 1)/k + \ldots,$$

$$\log\{k - \sqrt{(\zeta - k'^2)}\} = \log\{-\tfrac{1}{2}(\zeta - 1)/k + \ldots\} = \log(\zeta - 1) + P(\zeta - 1),$$

where $P(\zeta - 1)$ is a power series. Thus (3·5) gives

$$g(\zeta) = -2k \log(\zeta - 1) + P(\zeta - 1).$$

Using an asterisk to denote continuations after the circuit by ζ, (3·17) therefore gives

$${}^*g^{(p)}(\zeta) - g^{(p)}(\zeta) = -4k\pi i \zeta^p,$$

and (3·18) gives

$$\begin{aligned} {}^*g_m^{(p)}(\zeta) - g_m^{(p)}(\zeta) &= \zeta^{-m} \int_{\zeta}^{(1+)} z^{p+m-1} g(z)\, dz \\ &= -2k\zeta^{-m} \int_{\zeta}^{(1+)} z^{p+m-1} \log(z-1)\, dz \\ &= -2k\zeta^{-m} \left[\frac{z^{p+m}}{p+m} \log(z-1) - \int \frac{z^{p+m}\, dz}{(p+m)(z-1)} \right]_{\zeta}^{(1+)} \\ &= -4k\pi i \left(\frac{\zeta^p}{p+m} - \frac{\zeta^{-m}}{p+m} \right). \end{aligned}$$

Hence (3·16) gives

$$\frac{1}{4k\pi i}({}^*\Omega_p-\Omega_p) = -\omega(\tau)\,\zeta^p+\sum_{m=1}^{\infty}\frac{h_m(\tau T)^m F_m(\tau)}{p+m}(\zeta^p-\zeta^{-m}),$$

$$= -\left(\frac{\zeta}{T}\right)^p F_p(\tau)-\sum_{m=1}^{\infty}\frac{h_m}{p+m}\left(\frac{\tau T}{\zeta}\right)^m F_m(\tau), \tag{3·19}$$

by P.F.

In particular, since $F_0(\tau) = 1$,

$$\frac{1}{4k\pi i}({}^*\Omega_0-\Omega_0) = -1-\sum_{m=1}^{\infty}\frac{h_m}{m}\left(\frac{\tau T}{\zeta}\right)^m F_m(\tau), \tag{3·20}$$

and Ω_0 is essentially multiple-valued near $\zeta = 1$; it *and its derivatives* are multiple-valued. Since to a circuit of ζ round $\zeta = 1$, through the domains I, IV corresponds in the plane of the flow a circuit of large radius round the plane, the solution Ω_0 represents a flow with a line of discontinuity extending to infinity. Such a flow may well be physically significant (the line of discontinuity corresponding, say, to the wake behind an obstacle), but it is not the one which we are seeking.

4. CONSTRUCTION OF A SOLUTION WITH SINGLE-VALUED DERIVATIVES

Let

$$\bar{\zeta} = \frac{\tau^{\frac{1}{2}}T}{\tau_1^{\frac{1}{2}}T_1}\,\mathrm{e}^{-i\theta}, \tag{4·1}$$

so that

$$\zeta\bar{\zeta} = \tau T^2/\tau_1 T_1^2. \tag{4·2}$$

Then (3·19) can be written

$$\frac{1}{4k\pi i}({}^*\Omega_p-\Omega_p) = -\left(\frac{\zeta}{T}\right)^p F_p(\tau)-\sum_{m=1}^{\infty}\frac{h_m(\tau_1 T_1^2)^m}{p+m}\left(\frac{\bar{\zeta}}{T}\right)^m F_m(\tau). \tag{4·3}$$

Again, let $\bar{\Omega}_p$ be the conjugate solution to Ω_p:

$$\bar{\Omega}_p = k\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{\bar{\zeta}}{T}\right)^{n+p}F_{n+p}(\tau)+k\sum_{1}^{\infty}\frac{1}{\frac{1}{2}+n}\left(\frac{\bar{\zeta}}{T}\right)^{\frac{1}{2}+n+p}F_{\frac{1}{2}+n+p}(\tau)$$

$$+\sum_{0}^{\infty}\frac{q_n}{\frac{1}{2}-n}\left(\frac{\bar{\zeta}}{T}\right)^{\frac{1}{2}-n+p}F_{\frac{1}{2}-n+p}(\tau). \tag{4·4}$$

When ζ circles positively round $\zeta = 1$, $\bar{\zeta}$ circles negatively round $\bar{\zeta} = 1$, so

$$\frac{1}{4k\pi i}({}^*\bar{\Omega}_p-\bar{\Omega}_p) = \left(\frac{\bar{\zeta}}{T}\right)^p F_p(\tau)+\sum_{m=1}^{\infty}\frac{h_m(\tau_1 T_1^2)^m}{p+m}\left(\frac{\zeta}{T}\right)^m F_m(\tau). \tag{4·5}$$

From (3·20), (4·3) and (4·5) it follows that, if B_p, D_p are any constants, the solution

$$\Omega = \Omega_0+\sum_{p=1}^{\infty}h_p(\tau_1 T_1^2)^p\,(B_p\Omega_p+D_p\bar{\Omega}_p) \tag{4·6}$$

gives formally

$$\frac{1}{4k\pi i}({}^*\Omega-\Omega) = -1-\sum_{m=1}^{\infty}\frac{h_m(\tau_1 T_1^2)^m}{m}\left(\frac{\bar{\zeta}}{T}\right)^m F_m(\tau)$$

$$+\sum_{p=1}^{\infty}D_p h_p(\tau_1 T_1^2)^p\left\{\left(\frac{\bar{\zeta}}{T}\right)^p F_p(\tau)+\sum_{m=1}^{\infty}\frac{h_m(\tau_1 T_1^2)^m}{p+m}\left(\frac{\zeta}{T}\right)^m F_m(\tau)\right\}$$

$$-\sum_{p=1}^{\infty}B_p h_p(\tau_1 T_1^2)^p\left\{\left(\frac{\zeta}{T}\right)^p F_p(\tau)+\sum_{m=1}^{\infty}\frac{h_m(\tau_1 T_1^2)^m}{p+m}\left(\frac{\bar{\zeta}}{T}\right)^m F_m(\tau)\right\}. \tag{4·7}$$

Hence we shall have $${}^*\Omega - \Omega = -4k\pi i$$

provided

$$D_m - \sum_{p=1}^{\infty} \frac{B_p h_p(\tau_1 T_1^2)^p}{p+m} = \frac{1}{m}, \quad B_m - \sum_{p=1}^{\infty} \frac{D_p h_p(\tau_1 T_1^2)^p}{p+m} = 0 \quad (m = 1, 2, 3, \ldots). \tag{4·8}$$

From (4·8) by addition and subtraction we derive the equivalent system

$$(D_m + B_m) - \sum_{p=1}^{\infty} \frac{(D_p + B_p)\, h_p(\tau_1 T_1^2)^p}{p+m} = \frac{1}{m} \quad (m = 1, 2, 3, \ldots), \tag{4·9}$$

$$(D_m - B_m) + \sum_{p=1}^{\infty} \frac{(D_p - B_p)\, h_p(\tau_1 T_1^2)^p}{p+m} = \frac{1}{m} \quad (m = 1, 2, 3, \ldots). \tag{4·10}$$

Since, by (1·18), the hypothesis $\tau_1 < \tau_s$ gives $0 < \tau_1 T_1^2 < \delta^2$, and by (1·19)

$$h_p = 1/2\pi\delta^p\{1 + O(p^{-1})\},$$

we have $$0 < h_p(\tau_1 T_1^2)^p < A\xi^p, \tag{4·11}$$

where $0 < \xi < 1$ and A is a constant. The sum of the squares of the double series of coefficients in (4·9) or (4·10) is—apart from the units in the main diagonal—hence dominated by a series of the form $A \sum\limits_1^{\infty} \xi^{2p} \sum\limits_1^{\infty} (1+p)^{-2}$, so it is absolutely convergent; and since Σm^{-2} is convergent, the systems (4·9) and (4·10) are of Hilbert's 'fully continuous' type. Hence either (4·9) has a unique solution such that $\Sigma(B_p + D_p)^2$ converges, or the corresponding homogeneous system has such a solution; in the latter case a non-homogeneous system has in general no solution, and if a solution exists it is not unique. Actually we shall find an explicit solution for (4·9); this is unique for τ_1 sufficiently small, and probably for the full range $0 < \tau_1 < \tau_s$, but we shall not examine this question since the one solution is sufficient for our purposes. Similar remarks apply to (4·10).

To find a solution of (4·9) we write P.F. (1·13) in the form

$$\frac{F_\nu(\tau)}{T^\nu} + \sum_{p=1}^{\infty} \frac{h_p(\tau T^2)^p}{p+\nu} \frac{F_p(\tau)}{T^p} = \omega(\tau).$$

For $\nu = 0$ this is $$1 + \sum_{p=1}^{\infty} \frac{h_p(\tau T^2)^p}{p} \frac{F_p(\tau)}{T^p} = \omega(\tau),$$

and by subtraction, $$\frac{F_\nu(\tau)}{T^\nu} - \nu \sum_{p=1}^{\infty} \frac{h_p(\tau T^2)^p}{p+\nu} \frac{F_p(\tau)}{pT^p} = 1.$$

On dividing by ν we have, for $\nu = 1, 2, 3, \ldots$, precisely the form of (4·9), which has therefore the solution

$$D_m + B_m = F_m(\tau_1)/(mT_1^m). \tag{4·12}$$

To solve (4·10) we begin with (1·13) in the form

$$\tau^{\frac{1}{2}\nu} F_\nu(\tau) = (\tau^{\frac{1}{2}} T)^\nu \left\{ \omega(\tau) - \sum_{p=1}^{\infty} \frac{h_p(\tau^{\frac{1}{2}} T)^p \, \tau^{\frac{1}{2}p} F_p(\tau)}{p+\nu} \right\}.$$

Writing $$f_\nu(\tau) = \frac{d}{d\tau}\{\tau^{\frac{1}{2}\nu} F_\nu(\tau)\}, \quad \chi(\tau) = \frac{1}{\tau^{\frac{1}{2}} T} \frac{d}{d\tau}(\tau^{\frac{1}{2}} T), \tag{4·13}$$

this gives

$$f_\nu(\tau) = (\tau^{\frac{1}{2}}T)^\nu \{\omega'(\tau) + \nu\chi(\tau)\,\omega(\tau)\} - \sum_{p=1}^{\infty} h_p(\tau^{\frac{1}{2}}T)^{p+\nu}\,\tau^{\frac{1}{2}p}F_p(\tau)\,\chi(\tau) - \sum_{p=1}^{\infty}\frac{h_p(\tau^{\frac{1}{2}}T)^{p+\nu}f_p(\tau)}{p+\nu},$$

i.e.

$$\frac{f_\nu(\tau)}{(\tau^{\frac{1}{2}}T)^\nu} = \nu\chi(\tau)\,\omega(\tau) + \omega'(\tau) - \chi(\tau)\sum_{p=1}^{\infty} h_p(\tau T)^p F_p(\tau) - \sum_{p=1}^{\infty}\frac{h_p(\tau T^2)^p}{p+\nu}\frac{f_p(\tau)}{(\tau^{\frac{1}{2}}T)^p}$$

$$= \nu\chi(\tau)\,\omega(\tau) + \psi(\tau) - \sum_{p=1}^{\infty}\frac{h_p(\tau T^2)^p}{p+\nu}\frac{f_p(\tau)}{(\tau^{\frac{1}{2}}T)^p}, \quad \text{say.} \tag{4·14}$$

This result is valid for any ν that is not a negative integer. In particular for ν near 0 we easily find, from (1·7),

$$a_\nu = \beta + \tfrac{1}{2}\nu + O(\nu^2), \quad b_\nu = \tfrac{1}{2}\nu + O(\nu^2),$$

so that

$$F_\nu(\tau) = 1 + \tfrac{1}{2}\nu\left\{\beta\tau + \frac{\beta(\beta+1)}{2}\frac{\tau^2}{2!} + \frac{\beta(\beta+1)(\beta+2)}{3}\frac{\tau^3}{3!} + \ldots\right\} + O(\nu^2)$$

and

$$F'_\nu(\tau) = \tfrac{1}{2}\nu\left\{\beta + \frac{\beta(\beta+1)}{2!}\tau + \frac{\beta(\beta+1)(\beta+2)}{3!}\tau^2 + \ldots\right\} + O(\nu^2)$$

$$= \tfrac{1}{2}\nu\frac{(1-\tau)^{-\beta}-1}{\tau} + O(\nu^2).$$

Since (4·13) gives

$$f_\nu(\tau) = \tau^{\frac{1}{2}\nu}F'_\nu(\tau) + \tfrac{1}{2}\nu\tau^{\frac{1}{2}\nu-1}F_\nu(\tau),$$

we obtain

$$f_0(\tau) = 0, \quad \lim_{\nu\to 0}\frac{f_\nu(\tau)}{\nu} = \frac{(1-\tau)^{-\beta}}{2\tau} = \phi(\tau), \quad \text{say.} \tag{4·15}$$

Hence, subtracting from (4·14) the special case where $\nu = 0$, and dividing the result by ν, we obtain

$$\frac{f_\nu(\tau)}{\nu(\tau^{\frac{1}{2}}T)^\nu} = \chi(\tau)\,\omega(\tau) + \sum_{p=1}^{\infty}\frac{h_p(\tau T^2)^p}{p(p+\nu)}\frac{f_p(\tau)}{(\tau^{\frac{1}{2}}T)^p};$$

and for $\nu \to 0$ this gives, by (4·15),

$$\phi(\tau) = \chi(\tau)\,\omega(\tau) + \sum_{p=1}^{\infty}\frac{h_p(\tau T^2)^p}{p^2}\frac{f_p(\tau)}{(\tau^{\frac{1}{2}}T)^p}.$$

Subtracting this from the preceding,

$$\frac{f_\nu(\tau)}{\nu(\tau^{\frac{1}{2}}T)^\nu} + \nu\sum_{p=1}^{\infty}\frac{h_p(\tau T^2)^p}{p+\nu}\frac{f_p(\tau)}{p^2(\tau^{\frac{1}{2}}T)^p} = \phi(\tau),$$

and when this is divided by $\nu\phi(\tau)$ we obtain, for $\nu = 1, 2, 3, \ldots$, the form of (4·10). Hence (4·10) has the solution

$$D_m - B_m = \frac{f_m(\tau_1)}{m^2(\tau_1^{\frac{1}{2}}T_1)^m\,\phi(\tau_1)} = \frac{2\tau_1(1-\tau_1)^\beta}{m^2(\tau_1^{\frac{1}{2}}T_1)^m}\{\tau_1^{\frac{1}{2}m}F'_m(\tau_1) + \tfrac{1}{2}m\tau_1^{\frac{1}{2}m-1}F_m(\tau_1)\}$$

$$= \frac{(1-\tau_1)^\beta}{m^2T_1^m}\{2\tau_1F'_m(\tau_1) + mF_m(\tau_1)\}. \tag{4·16}$$

For a fixed τ between 0 and τ_s, it follows from P.F. that $F_m(\tau)/T^m$ and $F'_m(\tau)/mT^m$ remain bounded as $m \to \infty$. Hence (4·12) and (4·16) give

$$B_m = O(m^{-1}), \quad D_m = O(m^{-1}). \tag{4·17}$$

From (4·11) it then follows that the double series in (4·7) converge absolutely provided $|\zeta| < 1$, $|\bar{\zeta}| < 1$, and the formal rearrangement which leads to the conditions (4·8) is justified.

To complete the justification of the formal work we must consider not merely the increment formula (4·7) for the function Ω, but also the series (4·6) whereby this function is defined; since the increment formula presupposes the circling of ζ round the point 1, we must verify the convergence of (4·6) not merely for $k'^2 < |\zeta| < 1$, but also for ζ and $\bar{\zeta}$ traversing paths on which their moduli become somewhat greater than 1. For this purpose, let us first take $k'^2 < |\zeta| < 1$; then the series (3·15) defining Ω_p is dominated by

$$A\,|\zeta|^p \left\{ \sum_1^\infty |\zeta|^n + \sum_1^\infty |\zeta|^{\frac{1}{2}+n} + \sum_0^\infty |q_n| \cdot |\zeta|^{\frac{1}{2}-n} \right\},$$

where the series in $\{\ldots\}$ converges since (3·10) gives $|q_n| < Ak'^{2n}$. Hence, since the B_p, D_p are bounded, the series on the right of (4·6) is dominated by

$$A\left[1 + \sum_{p=1}^\infty h_p(\tau_1 T_1^2 |\zeta|)^p\right] \left\{ \sum_1^\infty |\zeta|^n + \sum_1^\infty |\zeta|^{\frac{1}{2}+n} + \sum_0^\infty |q_n| \cdot |\zeta|^{\frac{1}{2}-n} \right\};$$

and by (4·11) the series in $[\ldots]$ converges for $|\zeta| < \xi^{-1}$, where $\xi^{-1} > 1$.

Hence, for the range of τ corresponding to $k'^2 < |\zeta| < 1$, the series (4·6) for Ω converges absolutely, and we can substitute the series defining Ω_p, $\bar{\Omega}_p$ and then rearrange in any manner. In particular, we can collect all the hypergeometric functions of the same order, and thereby express Ω as a simple series of the standard form (1·21) + (1·22). Thence we can pass to an expression in terms of generating functions of ζ, $\bar{\zeta}$, say

$$\Omega = \{G(\zeta) + \bar{G}(\bar{\zeta})\}\,\omega(\tau) - \sum_{m=1}^\infty h_m(\tau T)^m F_m(\tau)\,\{G_m(\zeta) + \bar{G}_m(\bar{\zeta})\}. \tag{4·18}$$

This form is, of course, obtainable by substituting for the Ω_p, $\bar{\Omega}_p$ in (4·6) expansions such as (3·16), and then rearranging; the preceding argument justifies this process. The generating function for Ω_p is given in (3·17), whence we get, for the functions in (4·18),

$$G(\zeta) = g(\zeta) + \sum_{p=1}^\infty B_p h_p(\tau_1 T_1^2)^p g^{(p)}(\zeta) = g(\zeta)\left\{1 + \sum_p B_p h_p(\tau_1 T_1^2 \zeta)^p\right\},$$

$$\bar{G}(\bar{\zeta}) = \sum_{p=1}^\infty D_p h_p(\tau_1 T_1^2)^p g^{(p)}(\bar{\zeta}) = g(\bar{\zeta}) \sum_p D_p h_p(\tau_1 T_1^2 \bar{\zeta})^p.$$

The essential point now is that, in these expressions for $G(\zeta)$, $\bar{G}(\bar{\zeta})$, the series multiplying $g(\zeta)$ converge in $|\zeta| < \xi^{-1}$, a domain enclosing $\zeta = 1$, so the continuation of $g(\zeta)$ round $\zeta = 1$ leads immediately to that of $G(\zeta)$. Accordingly, by the general theory given in part I, the solution Ω can be correspondingly continued; the series for the continued function $^*\Omega$ can again be arranged in any manner we please, and the formula (4·7) is justified.

5. ANALYTIC CONTINUATION OF THE SOLUTION (4·6)

It is convenient to find the continuations of the separate terms of the series (4·6) and to deduce the continuation of Ω by addition; the justification of this formal process follows as at the end of the last section, the double series which arise being in every case absolutely convergent.

The function Ω_p is originally defined, in the domain I, by the series (3·15) or by the equivalent series (3·16). Its continuation along an assigned path into one of the other domains II, etc., is to be found by substituting in (3·16) the continuations of $g^{(p)}(\zeta)$, $g_m^{(p)}(\zeta)$, where, in I, $g^{(p)}(\zeta)$, $g_m^{(p)}(\zeta)$ are defined by the series (3·17) and (3·18). Since

$$g^{(p)}(\zeta) = \zeta^p g(\zeta), \tag{5·1}$$

the required continuations of $g^{(p)}(\zeta)$ can be read from (3·9). For $g_m^{(p)}(\zeta)$ the corresponding continuations are as follows; the proof is deferred to § 8 below:

$$\left.\begin{aligned}
g_m^{(p)}(\zeta) &= k\sum_{n=1}^{\infty}\frac{\zeta^{n+p}}{n(n+p+m)}+k\sum_{n=1}^{\infty}\frac{\zeta^{\frac{1}{2}+n+p}}{(\frac{1}{2}+n)(\frac{1}{2}+n+p+m)}+\sum_{n=0}^{\infty}\frac{q_n\zeta^{\frac{1}{2}-n+p}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)}, && \text{(I)}\\
g_m^{(p)}(\zeta) &= k\sum_{1}^{\infty}\frac{\zeta^{n+p}}{n(n+p+m)}-k\sum_{1}^{\infty}\frac{\zeta^{\frac{1}{2}+n+p}}{(\frac{1}{2}+n)(\frac{1}{2}+n+p+m)}-\sum_{0}^{\infty}\frac{q_n\zeta^{\frac{1}{2}-n+p}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)}, && \text{(I}'\text{)}\\
g_m^{(p)}(\zeta) &= -\frac{iC\zeta^p}{p+m}+\sum_{1}^{\infty}\frac{(k+ik'r_n)\,\zeta^{n+p}}{n(n+p+m)}+iK_{p+m}\zeta^{-m}, && \text{(II)}\\
g_m^{(p)}(\zeta) &= \frac{iC\zeta^p}{p+m}+\sum_{1}^{\infty}\frac{(k-ik'r_n)\,\zeta^{n+p}}{n(n+p+m)}-iK_{p+m}\zeta^{-m}, && \text{(III)}\\
g_m^{(p)}(\zeta) &= k\left(-\frac{2\pi i+\log\zeta}{p+m}+\frac{1}{(p+m)^2}\right)\zeta^p+k\sum_{1}^{\infty}{}'\,\frac{\zeta^{p-n}}{n(p+m-n)}\\
&\qquad -\sum_{0}^{\infty}\frac{s_n\zeta^{\frac{1}{2}-n+p}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)}+k\left(\frac{2\pi i+\log\zeta}{p+m}+\frac{1}{(p+m)^2}\right)\zeta^{-m}, && \text{(IV)}\\
g_m^{(p)}(\zeta) &= k\left(-\frac{\log\zeta}{p+m}+\frac{1}{(p+m)^2}\right)\zeta^p+k\sum_{1}^{\infty}{}'\,\frac{\zeta^{p-n}}{n(p+m-n)}\\
&\qquad +\sum_{0}^{\infty}\frac{s_n\zeta^{\frac{1}{2}-n+p}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)}+k\left(\frac{\log\zeta}{p+m}+\frac{1}{(p+m)^2}\right)\zeta^{-m}. && \text{(IV}'\text{)}
\end{aligned}\right\} \tag{5·2}$$

The coefficients q_n, r_n, s_n and the constant C are defined in (3·10); in II and III, K_r is the constant

$$K_r = \frac{\Gamma(\frac{3}{2})\,\Gamma(r)}{\Gamma(\frac{5}{2}+r)}\,k'^{3+2r}\,F(1,1+r,\tfrac{5}{2}+r,k'^2); \tag{5·3}$$

and in IV and IV′ Σ' indicates a summation in which the term with vanishing denominator is omitted.

For discussing the convergence of series involving the K_r it is useful to note that

$$F(1,1+r,\tfrac{5}{2}+r,k'^2) < F(1,1+r,\tfrac{5}{2}+r,1),$$

whence

$$0 < K_r < \frac{\Gamma(\frac{1}{2})\,\Gamma(r)}{\Gamma(\frac{3}{2}+r)}\,k'^{3+2r} < 2k'^{3+2r};$$

and, incidentally, C is the limit of rK_r as $r \to 0$.

Continuation into IV

On substituting in (3·16) the continuations for IV given by (3·9), (5·1) and (5·2) we obtain, on collecting like terms,

$$\begin{aligned}\Omega_p = &-k(2\pi i+\log\zeta)\,\zeta^p\left\{\omega(\tau)-\sum_{m=1}^{\infty}\frac{h_m(\tau T)^m F_m(\tau)}{p+m}\right\}\\ &+k\sum_{n=1}^{\infty}\frac{\zeta^{p-n}}{n}\left\{\omega(\tau)-\sum_{m=1}^{\infty}{}'\,\frac{h_m(\tau T)^m F_m(\tau)}{p-n+m}\right\}-k\zeta^p\sum_{m=1}^{\infty}\frac{h_m(\tau T)^m F_m(\tau)}{(p+m)^2}\\ &-\sum_{n=0}^{\infty}\frac{s_n\zeta^{\frac{1}{2}-n+p}}{\frac{1}{2}-n}\left\{\omega(\tau)-\sum_{m=1}^{\infty}\frac{h_m(\tau T)^m F_m(\tau)}{\frac{1}{2}-n+p+m}\right\}\\ &-k(2\pi i+\log\zeta)\sum_{m=1}^{\infty}\frac{h_m(\tau T/\zeta)^m F_m(\tau)}{p+m}-k\sum_{m=1}^{\infty}\frac{h_m(\tau T/\zeta)^m F_m(\tau)}{(p+m)^2}.\end{aligned}$$

As regards the first term in the second line, the m-summation involves no terms with vanishing denominator so long as $n \leqslant p$, and in this case we can substitute for $\{\dots\}$ from P.F.; also in the terms with $n > p$ we replace n by $n+p$. After substituting from P.F. in the first and third lines, we get

$$\begin{aligned}\Omega_p = &-k(2\pi i+\log\zeta)\left(\frac{\zeta}{T}\right)^p F_p(\tau)+k\sum_{n=1}^{\infty}\frac{\zeta^{-n}}{n+p}\left\{\omega(\tau)-\sum_{m=1}^{\infty}{}'\,\frac{h_m(\tau T)^m F_m(\tau)}{m-n}\right\}\\ &+k\sum_{n=1}^{p}\frac{1}{n}\left(\frac{\zeta}{T}\right)^{p-n}F_{p-n}(\tau)-k\zeta^p\sum_{m=1}^{\infty}\frac{h_m(\tau T)^m F_m(\tau)}{(p+m)^2}\\ &-\sum_{n=0}^{\infty}\frac{s_n}{\frac{1}{2}-n}\left(\frac{\zeta}{T}\right)^{\frac{1}{2}-n+p}F_{\frac{1}{2}-n+p}(\tau)\\ &-k(2\pi i+\log\zeta)\sum_{n=1}^{\infty}\frac{h_n(\tau T/\zeta)^n F_n(\tau)}{p+n}-k\sum_{n=1}^{\infty}\frac{h_n(\tau T/\zeta)^n F_n(\tau)}{(p+n)^2}.\end{aligned}$$

On reference to the expansions (2·4) and (2·6) for $L_n(\zeta)$, $M_n(\zeta)$ we get

$$\begin{aligned}\Omega_p = &-2\pi ik\left(\frac{\zeta}{T}\right)^p F_p(\tau)-2\pi ik\sum_{n=1}^{\infty}\frac{h_n(\tau T/\zeta)^n F_n(\tau)}{p+n}-kM_p(\zeta)+k\sum_{n=1}^{\infty}\frac{L_n(\zeta)}{n+p}\\ &+k\sum_{n=1}^{p}\frac{1}{n}\left(\frac{\zeta}{T}\right)^{p-n}F_{p-n}(\tau)-\sum_{n=0}^{\infty}\frac{s_n}{\frac{1}{2}-n}\left(\frac{\zeta}{T}\right)^{\frac{1}{2}-n+p}F_{\frac{1}{2}-n+p}(\tau)-k\sum_{n=1}^{\infty}\frac{h_n(\tau T/\zeta)^n F_n(\tau)}{(p+n)^2}.\end{aligned} \tag{5·4}$$

For the continuation of $\overline{\Omega}_p$ into IV we replace ζ by $\bar{\zeta}$, and also replace $-2\pi i$ by $2\pi i$, since ζ and $\bar{\zeta}$ circle in opposite senses round 1. Hence

$$\begin{aligned}\overline{\Omega}_p = &\,2\pi ik\left(\frac{\bar{\zeta}}{T}\right)^p F_p(\tau)+2\pi ik\sum_{n=1}^{\infty}\frac{h_n(\tau T/\bar{\zeta})^n F_n(\tau)}{p+n}-kM_p(\bar{\zeta})+k\sum_{n=1}^{\infty}\frac{L_n(\bar{\zeta})}{n+p}\\ &+k\sum_{n=1}^{p}\frac{1}{n}\left(\frac{\bar{\zeta}}{T}\right)^{p-n}F_{p-n}(\tau)-\sum_{n=0}^{\infty}\frac{s_n}{\frac{1}{2}-n}\left(\frac{\bar{\zeta}}{T}\right)^{\frac{1}{2}-n+p}F_{\frac{1}{2}-n+p}(\tau)-k\sum_{n=1}^{\infty}\frac{h_n(\tau T/\bar{\zeta})^n F_n(\tau)}{(p+n)^2}.\end{aligned} \tag{5·5}$$

The continuation of Ω is now deduced from

$$\Omega = \Omega_0+\sum_{p=1}^{\infty}\{B_p h_p(\tau_1 T_1^2)^p\,\Omega_p+D_p h_p(\tau_1 T_1^2)^p\,\overline{\Omega}_p\}. \tag{5·6}$$

The coefficient of $-2\pi ik$ is

$$1+\sum_{n=1}^{\infty}\frac{h_n(\tau T/\zeta)^n F_n(\tau)}{n}+\sum_{p=1}^{\infty}B_p h_p\left(\frac{\tau_1 T_1^2\zeta}{T}\right)^p F_p(\tau)+\sum_{p=1}^{\infty}B_p h_p(\tau_1 T_1^2)^p\sum_{n=1}^{\infty}\frac{h_n(\tau T/\zeta)^n F_n(\tau)}{p+n}$$
$$-\sum_{p=1}^{\infty}D_p h_p\left(\frac{\tau_1 T_1^2\bar{\zeta}}{T}\right)^p F_p(\tau)-\sum_{p=1}^{\infty}D_p h_p(\tau_1 T_1^2)^p\sum_{n=1}^{\infty}\frac{h_n(\tau T/\bar{\zeta})^n F_n(\tau)}{p+n}, \qquad (5\cdot7)$$

where by (4·2)
$$\frac{\tau_1 T_1^2\zeta}{T}=\frac{\tau T}{\bar{\zeta}},\quad \frac{\tau_1 T_1^2\bar{\zeta}}{T}=\frac{\tau T}{\zeta}.$$

From the relations (4·8) connecting the B_p, D_p it follows that the whole of the expression (5·7) cancels, except for the initial 1.

Again, the coefficients of $kL_n(\zeta)$, $kL_n(\bar{\zeta})$ are, for $n \geqslant 1$,

$$\frac{1}{n}+\sum_{p=1}^{\infty}\frac{B_p h_p(\tau_1 T_1^2)^p}{n+p},\quad \sum_{p=1}^{\infty}\frac{D_p h_p(\tau_1 T_1^2)^p}{n+p},$$

and by (4·8) these are respectively D_n, B_n. Hence the terms in Ω which involve the $L_n(\zeta)$, $L_n(\bar{\zeta})$, $M_n(\zeta)$, $M_n(\bar{\zeta})$ are

$$-kM_0(\zeta)+k\sum_{n=1}^{\infty}\{D_n L_n(\zeta)+B_n L_n(\bar{\zeta})-B_n h_n(\tau_1 T_1^2)^n M_n(\zeta)-D_n h_n(\tau_1 T_1^2)^n M_n(\bar{\zeta})\},$$

i.e. by (2·9)

$$-kM_0(\zeta)+k\sum_{n=1}^{\infty}\left\{D_n\left(\frac{\zeta}{T}\right)^{-n}+B_n\left(\frac{\bar{\zeta}}{T}\right)^{-n}\right\}\{\mathscr{F}_{-n}(\tau)+h_n\tau^n F_n(\tau)\log(\tau_1 T_1^2)\}.$$

On collecting the remaining terms in Ω, we have in all

$$\Omega=-2\pi ik-kM_0(\zeta)+k\sum_{n=1}^{\infty}\left\{D_n\left(\frac{\zeta}{T}\right)^{-n}+B_n\left(\frac{\bar{\zeta}}{T}\right)^{-n}\right\}\{\mathscr{F}_{-n}(\tau)+h_n\tau^n F_n(\tau)\log(\tau_1 T_1^2)\}$$
$$+k\sum_{p=1}^{\infty}h_p(\tau_1 T_1^2)^p\sum_{n=1}^{p}\frac{1}{n}\left\{B_p\left(\frac{\zeta}{T}\right)^{p-n}+D_p\left(\frac{\bar{\zeta}}{T}\right)^{p-n}\right\}F_{p-n}(\tau)$$
$$-\sum_{n=0}^{\infty}\frac{s_n}{\frac{1}{2}-n}\left[\left(\frac{\zeta}{T}\right)^{\frac{1}{2}-n}F_{\frac{1}{2}-n}(\tau)+\sum_{p=1}^{\infty}h_p(\tau_1 T_1^2)^p\left\{B_p\left(\frac{\zeta}{T}\right)^{\frac{1}{2}-n+p}+D_p\left(\frac{\bar{\zeta}}{T}\right)^{\frac{1}{2}-n+p}\right\}F_{\frac{1}{2}-n+p}(\tau)\right]$$
$$-k\sum_{n=1}^{\infty}h_n(\tau_1 T_1^2)^n\left(\frac{\bar{\zeta}}{T}\right)^n F_n(\tau)\left\{\frac{1}{n^2}+\sum_{p=1}^{\infty}\frac{B_p h_p(\tau_1 T_1^2)^p}{(p+n)^2}\right\}$$
$$-k\sum_{n=1}^{\infty}h_n(\tau_1 T_1^2)^n\left(\frac{\zeta}{T}\right)^n F_n(\tau)\sum_{p=1}^{\infty}\frac{D_p h_p(\tau_1 T_1^2)^p}{(p+n)^2}. \quad \text{(IV)} \qquad (5\cdot8)$$

The cancellation of the variable part of the coefficient of $2\pi i$ is, of course, a verification that, in I + IV, Ω has single-valued derivatives. If the continuation had been made in the opposite sense from I into IV, we should have obtained a formula for Ω which differs only from (5·8) in having $2\pi ik$ in place of $-2\pi ik$ for the first term.

By the general theory of analytic continuation given in part I we know that the series (5·8) must converge for $\tau_1<\tau<\tau_s$; and we can verify from (1·17)–(1·19), (2·12) and (4·17) that it converges actually for $\tau_1<\tau<1$. We note that as τ increases from τ_1 through τ_s to 1, $|\zeta|$ increases from 1 to $\delta(\tau_1^{\frac{1}{2}}T_1)^{-1}$ and then remains constant. The functions $\mathscr{F}_{-n}(\tau)$, $h_n\tau^n F_n(\tau)\log(\tau_1 T_1^2)$ are both of order $n^{\frac{1}{2}}|T|^{-n}$ when $\tau\geqslant\tau_s$, while for $\tau < \tau_s$ the former is $O(T^{-n})$ and the latter is smaller. Hence, since B_n, D_n are bounded, the series in the first line of (5·8) is dominated by $A\Sigma n^{\frac{1}{6}} |\zeta|^{-n}$, and converges for $\tau > \tau_s$.

In the series in the second line, $p-n$ is positive in every term, and the inner (finite) sum is dominated by $A|\zeta|^{p-1} \sum\limits_{n=1}^{p} n^{-\frac{5}{6}}$, i.e. by $Ap^{\frac{1}{6}}|\zeta|^p$, or by $Ap^{\frac{1}{6}}$, according as $|\zeta| \geqslant 1$ or $|\zeta| < 1$. The series is hence dominated by

$$A\sum_{p=1}^{\infty} \frac{p^{\frac{1}{6}} |\zeta|^p (\tau_1 T_1^{\frac{2}{3}})^p}{\delta^{2p}}, \quad \text{or by} \quad A\sum_{p=1}^{\infty} \frac{p^{\frac{1}{6}} (\tau_1 T_1^{\frac{2}{3}})^p}{\delta^{2p}},$$

and in the least favourable case, when $|\zeta| = \delta(\tau_1^{\frac{1}{2}} T_1)^{-1}$, the dominant is $A\Sigma p^{\frac{1}{6}}(\tau_1^{\frac{1}{2}} T_1/\delta)^p$, which converges. Thus the second line converges for $0 \leqslant \tau < 1$.

For the series in the third line, the inner sum is dominated by

$$A|\zeta|^{\frac{1}{2}-n} \sum_{p=1}^{\infty} \frac{p^{\frac{1}{6}} |\zeta|^p (\tau_1 T_1^{\frac{2}{3}})^p}{\delta^{2p}},$$

which has just been shown to converge. Hence, since the s_n are bounded, the complete series is dominated by $A\Sigma |\zeta|^{\frac{1}{2}-n} n^{-\frac{5}{6}}$, which converges for $\tau > \tau_1$.

For the remaining series, the coefficients

$$\frac{1}{n^2} + \sum_{p=1}^{\infty} \frac{B_p h_p (\tau_1 T_1^{\frac{2}{3}})^p}{(p+n)^2}, \qquad \sum_{p=1}^{\infty} \frac{D_p h_p (\tau_1 T_1^{\frac{2}{3}})^p}{(p+n)^2}$$

are $O(1/n^2)$, and the series converge for $0 \leqslant \tau < 1$.

Continuation into I′, IV′

The expansions of $g^{(p)}(\zeta)$, $g_m^{(p)}(\zeta)$ in I′ are derived from those in I by reversing the signs of the terms of half-integral order. Hence the expression for Ω in I′ is derived from that in I by making the corresponding change, and from (3·15) and (4·6) we have in I′

$$\begin{aligned}\Omega = k\sum_{n=1}^{\infty} \frac{1}{n}\left[\left(\frac{\zeta}{T}\right)^{n} F_n(\tau) + \sum_{p=1}^{\infty} h_p(\tau_1 T_1^{\frac{2}{3}})^p \left\{B_p\left(\frac{\zeta}{T}\right)^{n+p} + D_p\left(\frac{\bar{\zeta}}{T}\right)^{n+p}\right\} F_{n+p}(\tau)\right] \\ - k\sum_{n=1}^{\infty} \frac{1}{\frac{1}{2}+n}\left[\left(\frac{\zeta}{T}\right)^{\frac{1}{2}+n} F_{\frac{1}{2}+n}(\tau) + \sum_{p=1}^{\infty} h_p(\tau_1 T_1^{\frac{2}{3}})^p \left\{B_p\left(\frac{\zeta}{T}\right)^{\frac{1}{2}+n+p} + D_p\left(\frac{\bar{\zeta}}{T}\right)^{\frac{1}{2}+n+p}\right\} F_{\frac{1}{2}+n+p}(\tau)\right] \\ - \sum_{n=0}^{\infty} \frac{q_n}{\frac{1}{2}-n}\left[\left(\frac{\zeta}{T}\right)^{\frac{1}{2}-n} F_{\frac{1}{2}-n}(\tau) + \sum_{p=1}^{\infty} h_p(\tau_1 T_1^{\frac{2}{3}})^p \left\{B_p\left(\frac{\zeta}{T}\right)^{\frac{1}{2}-n+p} + D_p\left(\frac{\bar{\zeta}}{T}\right)^{\frac{1}{2}-n+p}\right\} F_{\frac{1}{2}-n+p}(\tau)\right]. \quad (\mathrm{I}')\end{aligned} \tag{5·9}$$

Similarly, the expansion for IV′ is derived from that for IV, viz. (5·8), by reversing the sign before the series (in the third line) of terms of half-integral order, and also omitting the term $-2\pi ik$. The absence of the latter term reflects the fact that Ω is non-singular at the point $\zeta = 1$ on the common boundary of I′ and IV′.

Continuations into II, III

On substituting in (3·16) the appropriate expansions of $g^{(p)}(\zeta)$, $= \zeta^p g(\zeta)$, and $g_m^{(p)}(\zeta)$, as given in (3·9) and (5·2), we obtain for II, after collecting like terms,

$$\begin{aligned}\Omega_p &= \sum_{n=1}^{\infty} \frac{k+ik'r_n}{n} \zeta^{n+p} \left\{ \omega(\tau) - \sum_{m=1}^{\infty} \frac{h_m(\tau T)^m F_m(\tau)}{n+p+m} \right\} \\ &\quad -iC\zeta^p \left\{ \omega(\tau) - \sum_{m=1}^{\infty} \frac{h_m(\tau T)^m F_m(\tau)}{p+m} \right\} - i \sum_{m=1}^{\infty} K_{p+m} h_m \left(\frac{\tau T}{\zeta}\right)^m F_m(\tau) \\ &= \sum_{n=1}^{\infty} \frac{k+ik'r_n}{n} \left(\frac{\zeta}{T}\right)^{n+p} F_{n+p}(\tau) - iC\left(\frac{\zeta}{T}\right)^p F_p(\tau) - i \sum_{n=1}^{\infty} K_{p+n} h_n (\tau_1 T_1^2)^n \left(\frac{\bar{\zeta}}{T}\right)^n F_n(\tau). \quad \text{(II)}\end{aligned} \tag{5·10}$$

For $\bar{\Omega}_p$ we replace ζ by $\bar{\zeta}$ and i by $-i$:

$$\bar{\Omega}_p = \sum_{n=1}^{\infty} \frac{k-ik'r_n}{n} \left(\frac{\bar{\zeta}}{T}\right)^{n+p} F_{n+p}(\tau) + iC\left(\frac{\bar{\zeta}}{T}\right)^p F_p(\tau) + i \sum_{n=1}^{\infty} K_{p+n} h_n (\tau_1 T_1^2)^n \left(\frac{\zeta}{T}\right)^n F_n(\tau). \quad \text{(II)} \tag{5·11}$$

To obtain Ω we substitute these in (15·6); the full formula does not simplify, and is not worth writing *in extenso*.

For III we have the same formulae with i replaced by $-i$.

6. REMARKS

(i) Our solution Ω has now been continued through the domain $0 \leqslant \tau < 1$, $-\pi \leqslant \theta \leqslant \pi$. It has two branches which permute about the branch point $\tau = \tau'$, $\theta = 0$, where τ' is determined by the equation $\zeta = k'^2$, i.e. $\sqrt{\tau'}\, T(\tau') = k'^2 \sqrt{\tau_1}\, T(\tau_1)$. One of these branches has a simple logarithmic point at $\tau = \tau_1$, $\theta = 0$, so that its derivatives with respect to τ, θ are single valued near this point and infinite at it; the other branch is regular at this point.

From the discussion at the beginning of § 3, therefore, we see that our solution has the correct functional character for the specification of flow, with circulation, past a cylinder. We can verify this by examining its approximate form when τ_1 is small and τ is restricted to be comparable with τ_1. Then the only significant terms are those involving the quotient τ/τ_1; at every other occurrence of τ or τ_1, we can replace them by 0. On this approximation, all the hypergeometric functions become equal to 1, and ζ/T become $\sqrt{\tau}\, e^{i\theta}/\sqrt{\tau_1} = q\, e^{i\theta}/q_1$, since T, T_1 become 1; also Ω reduces to Ω_0, since the series $\sum_p \{B_p h_p (\tau_1 T_1^2)^p \Omega_p + D_p h_p (\tau_1 T_1^2)^p \bar{\Omega}_p\}$ vanishes with τ_1. From (3·14) we see then that Ω reduces to $g(q\, e^{i\theta}/q_1)$, and since this specifies incompressible flow, with circulation, past a cylinder, we verify what is required.

(ii) The physical solution is obtained from our Ω by multiplying by an arbitrary pure-imaginary scale-factor and then adding the conjugate. Hence it is

$$\Omega_{\text{phys.}} = A\left\{ i(\Omega_0 - \bar{\Omega}_0) + \sum_{p=1}^{\infty} h_p (\tau_1 T_1^2)^p (B_p - D_p)\, i(\Omega_p - \bar{\Omega}_p) \right\}; \tag{6·1}$$

the value of the real constant $B_p - D_p$ is given in (4·16).

The position co-ordinates x, y are thence given in terms of τ, θ by differentiation as in (1·3). Separate series are of course required for the domains I, ..., IV′; their derivation is merely a matter of routine, and they will not here be given.

We can easily verify that the solution (6·1) gives a flow in the xy-plane whose circulation round any circuit of large radius, with centre at $x = y = 0$, is exactly $8\pi kA$. In fact, the circulation is

$$\begin{aligned}\int(u\,dx+v\,dy) &= [ux+vy]-\int(x\,du+y\,dv)\\ &= 0-\int\left(\frac{\partial\Omega}{\partial u}du+\frac{\partial\Omega}{\partial v}dv\right) = -[\Omega],\end{aligned}$$

where $[\Omega]$ is the increment of Ω on the circuit. To a large circuit in the xy-plane corresponds a small circuit of ζ about $\zeta = 1$, and for this we have found the increment of Ω to be $-4k\pi i$. Hence the increment of $\Omega-\bar{\Omega}$ is $-8k\pi i$, and the increment of $\Omega_{\text{phys.}}$ is $8\pi kA$.

It may be remarked that our solution remains regular at the stagnation points, given by $\zeta = \bar{\zeta} = 0$ in II and III; that is, the series (5·10) and (5·11), and the series for Ω which thence follows, involve only terms of positive integral orders.

(iii) Our solution Ω involves the two essential parameters τ_1, giving the speed of the flow at infinity (which must be subsonic since the restriction $\tau_1 < \tau_s$ has been essential), and k, giving the circulation; the restriction $0 < k < 1$ is not essential, and by similar work we can find the formulae in the cases where $k > 1$. We can obtain other solutions, of the same physical character, by adding to our Ω any solution which introduces no new singularities, e.g. (1) a series of the form

$$\sum_{n=1}^{\infty}(A_n\,\mathrm{e}^{ni\theta}+\bar{A}_n\,\mathrm{e}^{-ni\theta})\,\tau^{\frac{1}{2}n}F_n(\tau), \tag{6·2}$$

which converges for $0 \leqslant \tau < 1$, or (2) a solution whose generating function is

$$A\surd(\zeta-k'^2),$$

as developed in part I. We have encountered series of the sort (1) in (5·8), second, fourth and fifth lines; and the last series in (5·10) has also this character.

It is easy to prove that our solution Ω gives a flow having the line $x = 0$ as an axis of symmetry; this is because the coefficients in the developments for I, I′, IV, IV′ are real. If we add a series of the form (6·2) whose coefficients are not all real we destroy this symmetry.

(iv) The formula (3·20) gives $({}^*\Omega_0-\Omega_0)/4k\pi i$ as the sum of a 'principal term' -1 and a series of 'subsidiary terms' $\Sigma h_m(\tau_1 T_1^2)^m(\bar{\zeta}/T)^m F_m(\tau)/m$. We cancelled these subsidiary terms by adding to Ω_0 a series of other solutions Ω_p, $\bar{\Omega}_p$; the reason we were successful in this is that the principal term in ${}^*\bar{\Omega}_p-\bar{\Omega}_p$ has the form of the pth subsidiary term in ${}^*\Omega_0-\Omega_0$; and, as is easily seen, the essential property of $\bar{\Omega}_p$ for this purpose is that its generating function has near $\bar{\zeta} = 1$ the principal part $A\bar{\zeta}^p\log(\bar{\zeta}-1)$, where A is constant. Any function $\bar{\Omega}_p$ having this character would serve equally well so far as the neighbourhood of $\bar{\zeta} = 1$ is concerned; but we require also that the final solution Ω should have no other singularities than those of Ω_0, so that for instance $\overline{\Omega}_p$ must be double valued, with its second branch regular at $\bar{\zeta} = 1$. This additional requirement is satisfied by our choice of $\overline{\Omega}_p$ as the solution whose generating function is $\bar{\zeta}^p g(\bar{\zeta})$, but it could have been satisfied by other choices of generating function, e.g. $k\bar{\zeta}^p \log g'(\bar{\zeta})$. The latter choice gives a solution which reduces, when $k \to 0$, to that found in part I for flow without circulation; the choice which we have adopted gives a solution which does not so reduce: for $k = 0$ it gives a flow without circulation round a cylinder, but the shapes of the cylinders are not the same in the two cases, unless τ_1 is infinitesimal.

7. The stream function

We begin by choosing a series for $i\psi$ which has in the domain I the correct relation to Ω, according to (1·11) and (1·12). From this we find the series for $i\psi$ in the other domains by continuation.

Accordingly we start from†

$$i\psi = \psi_0 + \sum_{p=1}^{\infty} h_p(\tau_1 T_1^2)^p \,(B_p \psi_p - D_p \overline{\psi}_p), \tag{7·1}$$

where in I

$$\begin{aligned}\psi_p = k\sum_{n=1}^{\infty} \frac{p+n-1}{n}\left(\frac{\zeta}{T}\right)^{p+n} \tilde{F}_{p+n}(\tau) + k\sum_{n=1}^{\infty} \frac{p+n-\frac{1}{2}}{\frac{1}{2}+n}\left(\frac{\zeta}{T}\right)^{\frac{1}{2}+p+n} \tilde{F}_{\frac{1}{2}+p+n}(\tau) \\ + \sum_{n=0}^{\infty} \frac{(p-n-\frac{1}{2})\,q_n}{\frac{1}{2}-n}\left(\frac{\zeta}{T}\right)^{\frac{1}{2}+p-n} \tilde{F}_{\frac{1}{2}+p-n}(\tau), \quad \text{(I)}\end{aligned} \tag{7·2}$$

and $\overline{\psi}_p$ is the same function of $\bar{\zeta}$. From P.F. this is equivalent to

$$\psi_p = \tilde{g}^{(p)}(\zeta)\,\tilde{\omega}(\tau) - \sum_{m=1}^{\infty} \tilde{h}_m(\tau T)^m \tilde{F}_m(\tau)\,\tilde{g}_m^{(p)}(\zeta), \tag{7·3}$$

where

$$\left.\begin{aligned}\tilde{g}^{(p)}(\zeta) &= k\sum_{n=1}^{\infty} \frac{p+n-1}{n}\zeta^{p+n} + k\sum_{1}^{\infty} \frac{p+n-\frac{1}{2}}{\frac{1}{2}+n}\zeta^{\frac{1}{2}+p+n} + \sum_{0}^{\infty} \frac{(p-n-\frac{1}{2})\,q_n}{\frac{1}{2}-n}\zeta^{\frac{1}{2}+p-n}, \\ \tilde{g}_m^{(p)}(\zeta) &= k\sum_{1}^{\infty} \frac{(p+n-1)\,\zeta^{p+n}}{n(p+n+m)} + k\sum_{1}^{\infty} \frac{(p+n-\frac{1}{2})\,\zeta^{\frac{1}{2}+p+n}}{(\frac{1}{2}+n)\,(\frac{1}{2}+p+n+m)} + \sum_{0}^{\infty} \frac{(p-n-\frac{1}{2})\,q_n \zeta^{\frac{1}{2}+p-n}}{(\frac{1}{2}-n)\,(\frac{1}{2}-n+p+m)}.\end{aligned}\right\} \tag{7·4}$$

Hence, from (3·17) and (3·18),

$$\tilde{g}^{(p)}(\zeta) = \zeta\frac{d}{d\zeta}g^{(p)}(\zeta) - g^{(p)}(\zeta), \quad \tilde{g}_m^{(p)}(\zeta) = \zeta\frac{d}{d\zeta}g_m^{(p)}(\zeta) - g_m^{(p)}(\zeta), \tag{7·5}$$

and thence from the known continuations of $g^{(p)}(\zeta)$, $g_m^{(p)}(\zeta)$ we can deduce those of $\tilde{g}^{(p)}(\zeta)$, $\tilde{g}_m^{(p)}(\zeta)$.

Now in each of the domains I′, II, III we see from (3·9) and (5·2) that the expansions of the g-functions are related as in the scheme

$$g^{(p)}(\zeta) = \sum_{\nu} C_\nu \zeta^\nu, \quad g_m^{(p)}(\zeta) = \sum_{\nu} \frac{C_\nu \zeta^\nu}{\nu+m} + A_m \zeta^{-m}, \tag{7·6}$$

---

† We have chosen the notation ψ_p, rather than $i\psi_p$, for the series which make up $i\psi$, because the coefficients in these series, for I, are real.

where ν runs over a set of values which includes no negative integers, and A_m is a constant. By (7·5) these give

$$\tilde{g}^{(p)}(\zeta) = \sum_{\nu} (\nu-1)\, C_\nu \zeta^\nu, \quad \tilde{g}^{(p)}_m(\zeta) = \sum_{\nu} \frac{(\nu-1)\, C_\nu \zeta^\nu}{\nu+m} - (m+1)\, A_m \zeta^{-m},$$

and by substitution in (7·3) and appeal to P.F. we get

$$\psi_p = \sum_{\nu} (\nu-1)\, C_\nu \left(\frac{\zeta}{T}\right)^{\nu} \tilde{F}_\nu(\tau) + \sum_{m=1}^{\infty} A_m \tilde{h}_m (m+1) \left(\frac{\tau T}{\zeta}\right)^m \tilde{F}_m(\tau). \tag{7·7}$$

Similarly, (7·6) gives

$$\Omega_p = \sum_{\nu} C_\nu \left(\frac{\zeta}{T}\right)^{\nu} F_\nu(\tau) - \sum_{m=1}^{\infty} A_m h_m \left(\frac{\tau T}{\zeta}\right)^m F_m(\tau). \tag{7·8}$$

Since $\quad (\tau T/\zeta)^m = (\tau_1 T_1^2)^m (\bar{\zeta}/T)^m \quad$ and $\quad (m+1)\,\tilde{h}_m = (m-1)\, h_m,$

we verify that the solutions (7·8) and (7·7) are related as in the scheme (1·11) and (1·12)—as, indeed, they must be; but we prove also something more, viz. that when $g^{(p)}(\zeta)$, $g^{(p)}_m(\zeta)$ are related as in (7·6), the additive constant which remained undetermined in (1·12) is zero.

Hence we can read off the formulae for ψ_p in I′, II, III from those for Ω_p; viz. in I′, a series got from (7·2) by reversing the signs of the terms of half-integral order; in II,

$$\psi_p = \sum_{n=1}^{\infty} \frac{(k+ik'r_n)(n+p-1)}{n} \left(\frac{\zeta}{T}\right)^{n+p} \tilde{F}_{n+p}(\tau) - i(p-1)\, C \left(\frac{\zeta}{T}\right)^{p} \tilde{F}_p(\tau)$$
$$+ i \sum_{n=1}^{\infty} (n-1)\, K_{p+n} h_n (\tau_1 T_1^2)^n \left(\frac{\bar{\zeta}}{T}\right)^{n} \tilde{F}_n(\tau); \quad \text{(II)} \tag{7·9}$$

and in III the same series with i replaced by $-i$.

To get $\bar{\psi}_p$ from ψ_p we interchange ζ, $\bar{\zeta}$ and i, $-i$, and by substitution in (7·1) we get the series for $i\psi$ in I′, II, III.

For IV, we begin from the appropriate expansions of $g^{(p)}(\zeta)$, $g^{(p)}_m(\zeta)$, viz.

$$g^{(p)}(\zeta) = -k(2\pi i + \log\zeta)\,\zeta^p + k \sum_{1}^{\infty} \frac{\zeta^{p-n}}{n} - \sum_{0}^{\infty} \frac{s_n}{\frac{1}{2}-n} \zeta^{\frac{1}{2}+p-n},$$

$$g^{(p)}_m(\zeta) = -k\left(\frac{2\pi i + \log\zeta}{p+m} - \frac{1}{(p+m)^2}\right)\zeta^p + k \sum_{1}^{\infty}{}' \frac{\zeta^{p-n}}{n(p+m-n)}$$
$$- \sum_{0}^{\infty} \frac{s_n \zeta^{\frac{1}{2}+p-n}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)} + k\left(\frac{2\pi i + \log\zeta}{p+m} + \frac{1}{(p+m)^2}\right)\zeta^{-m}.$$

From (7·5) we deduce

$$\tilde{g}^{(p)}(\zeta) = -k(p-1)(2\pi i + \log\zeta)\,\zeta^p + k \sum_{1}^{\infty} \frac{(p-n-1)\,\zeta^{p-n}}{n} - \sum_{0}^{\infty} \frac{(p-n-\frac{1}{2})\, s_n \zeta^{\frac{1}{2}+p-n}}{\frac{1}{2}-n} - k\zeta^p,$$

$$\tilde{g}^{(p)}_m(\zeta) = -k(p-1)\left(\frac{2\pi i + \log\zeta}{p+m} - \frac{1}{(p+m)^2}\right)\zeta^p + k \sum_{1}^{\infty} \frac{(p-n-1)\,\zeta^{p-n}}{n(p+m-n)} - \sum_{0}^{\infty} \frac{(p-n-\frac{1}{2})\, s_n \zeta^{\frac{1}{2}+p-n}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)}$$
$$- k(m+1)\left(\frac{2\pi i + \log\zeta}{p+m} + \frac{1}{(p+m)^2}\right)\zeta^{-m} - \frac{k}{p+m}\zeta^p + \frac{k}{p+m}\zeta^{-m}.$$

On substitution in (7·3) and reduction as for Ω_p we get

$$\begin{aligned}\psi_p = &-2\pi ki(p-1)\left(\frac{\zeta}{T}\right)^p \tilde{F}_p(\tau) + 2\pi ki \sum_{n=1}^{\infty} \frac{(n+1)\,\tilde{h}_n}{p+n}\left(\frac{\tau T}{\zeta}\right)^n \tilde{F}_n(\tau) - k(p-1)\,\tilde{M}_p(\zeta) \\ &-k\sum_{n=1}^{\infty}\frac{n+1}{p+n}\tilde{L}_n(\zeta) + k\sum_{n=1}^{p}\frac{p-n-1}{n}\left(\frac{\zeta}{T}\right)^{p-n}\tilde{F}_{p-n}(\tau) - \sum_{n=0}^{\infty}\frac{(p-n-\frac{1}{2})\,s_n}{\frac{1}{2}-n}\left(\frac{\zeta}{T}\right)^{\frac{1}{2}+p-n}\tilde{F}_{\frac{1}{2}+p-n}(\tau) \\ &+k\sum_{n=1}^{\infty}\frac{(n+1)\,\tilde{h}_n}{(p+n)^2}\left(\frac{\tau T}{\zeta}\right)^n \tilde{F}_n(\tau) - k\left(\frac{\zeta}{T}\right)^p \tilde{F}_p(\tau) - \sum_{n=1}^{\infty}\frac{k\tilde{h}_n}{p+n}\left(\frac{\tau T}{\zeta}\right)^n \tilde{F}_n(\tau); \quad \text{(IV)}\end{aligned} \tag{7·10}$$

and $\overline{\psi}_p$ is the conjugate of this:

$$\begin{aligned}\overline{\psi}_p = &\;2\pi ki(p-1)\left(\frac{\bar{\zeta}}{T}\right)^p \tilde{F}_p(\tau) - 2\pi ki \sum_{n=1}^{\infty} \frac{(n+1)\,\tilde{h}_n}{p+n}\left(\frac{\tau T}{\bar{\zeta}}\right)^n \tilde{F}_n(\tau) - k(p-1)\,\tilde{M}_p(\bar{\zeta}) \\ &-k\sum_{n=1}^{\infty}\frac{n+1}{p+n}\tilde{L}_n(\bar{\zeta}) + k\sum_{n=1}^{p}\frac{p-n-1}{n}\left(\frac{\bar{\zeta}}{T}\right)^{p-n}\tilde{F}_{p-n}(\tau) - \sum_{n=0}^{\infty}\frac{(p-n-\frac{1}{2})\,s_n}{\frac{1}{2}-n}\left(\frac{\bar{\zeta}}{T}\right)^{\frac{1}{2}+p-n}\tilde{F}_{\frac{1}{2}+p-n}(\tau) \\ &+k\sum_{n=1}^{\infty}\frac{(n+1)\,\tilde{h}_n}{(p+n)^2}\left(\frac{\tau T}{\bar{\zeta}}\right)^n \tilde{F}_n(\tau) - k\left(\frac{\bar{\zeta}}{T}\right)^p \tilde{F}_p(\tau) - \sum_{n=1}^{\infty}\frac{k\tilde{h}_n}{p+n}\left(\frac{\tau T}{\bar{\zeta}}\right)^n \tilde{F}_n(\tau). \quad \text{(IV)}\end{aligned} \tag{7·11}$$

From these we obtain $i\psi$ by substitution in (7·1), viz.

$$i\psi = \psi_0 + \sum_{p=1}^{\infty} h_p(\tau_1 T_1^2)^p\,(B_p\psi_p - D_p\overline{\psi}_p). \tag{7·12}$$

Observing that $\qquad (n+1)\,\tilde{h}_n = (n-1)\,h_n,$

we get for the coefficient of $2\pi ki$

$$\begin{aligned}1 &+ \sum_{n=1}^{\infty}\frac{(n-1)\,h_n}{n}\left(\frac{\tau T}{\zeta}\right)^n \tilde{F}_n(\tau) - \sum_{p=1}^{\infty}(p-1)\,B_p h_p\left(\frac{\tau_1 T_1^2 \zeta}{T}\right)^p \tilde{F}_p(\tau) \\ &+ \sum_{p=1}^{\infty} B_p h_p(\tau_1 T_1^2)^p \sum_{n=1}^{\infty}\frac{(n-1)\,h_n}{p+n}\left(\frac{\tau T}{\zeta}\right)^n - \sum_{p=1}^{\infty}(p-1)\,D_p h_p\left(\frac{\tau_1 T_1^2 \bar{\zeta}}{T}\right)^p \tilde{F}_p(\tau) \\ &+ \sum_{p=1}^{\infty} D_p h_p(\tau_1 T_1^2)^p \sum_{n=1}^{\infty}\frac{(n-1)\,h_n}{p+n}\left(\frac{\tau T}{\bar{\zeta}}\right)^n \tilde{F}_n(\tau),\end{aligned}$$

an expression which reduces to 1 in the same way as (5·7) does. Again, the terms involving the functions $\tilde{L}$, $\tilde{M}$ reduce, on use of (4·8), to

$$\begin{aligned}k\tilde{M}_0(\zeta) - k\sum_{n=1}^{\infty}\{(n+1)\,D_n\tilde{L}_n(\zeta) - (n+1)\,B_n\tilde{L}_n(\bar{\zeta}) + (n-1)\,h_n(\tau_1 T_1^2)^n\,B_n\tilde{M}_n(\zeta) \\ - (n-1)\,h_n(\tau_1 T_1^2)^n\,D_n\tilde{M}_n(\bar{\zeta})\},\end{aligned}$$

and putting $(n-1)\,h_n = (n+1)\,\tilde{h}_n$ this becomes, by the analogue of (2·9),

$$k\tilde{M}_0(\zeta) - k\sum_{n=1}^{\infty}(n+1)\left\{D_n\left(\frac{\zeta}{T}\right)^{-n} - B_n\left(\frac{\bar{\zeta}}{T}\right)^{-n}\right\}\{\mathscr{F}_{-n}(\tau) + \tilde{h}_n\tau^n\tilde{F}_n(\tau)\log(\tau_1 T_1^2)\}.$$

Similarly, the last two terms in (7·10) and (7·11) give the contribution

$$-k - k\sum_{n=1}^{\infty}\left\{B_n h_n(\tau_1 T_1^2)^n\left(\frac{\zeta}{T}\right)^n + D_n\tilde{h}_n\left(\frac{\tau T}{\zeta}\right)^n - D_n h_n(\tau_1 T_1^2)^n\left(\frac{\bar{\zeta}}{T}\right)^n - B_n\tilde{h}_n\left(\frac{\tau T}{\bar{\zeta}}\right)^n\right\}\tilde{F}_n(\tau),$$

which reduces to

$$-k+k\sum_{n=1}^{\infty}\frac{2h_n}{n+1}\left\{D_n\left(\frac{\zeta}{T}\right)^{-n}-B_n\left(\frac{\bar{\zeta}}{T}\right)^{-n}\right\}\tau^n\tilde{F}_n(\tau),$$

since $\zeta\bar{\zeta}=\tau T^2/(\tau_1 T_1^2)$ and $\tilde{h}_n=(n-1)h_n/(n+1)$.

On collecting the remaining terms, (7·12) becomes

$$\begin{aligned}
i\psi = {} & 2\pi ki-k+k\tilde{M}_0(\zeta)-k\sum_{n=1}^{\infty}(n+1)\left\{D_n\left(\frac{\zeta}{T}\right)^{-n}-B_n\left(\frac{\bar{\zeta}}{T}\right)^{-n}\right\}\\
& \times\left\{\mathscr{F}_{-n}(\tau)+\tilde{h}_n\tau^n\tilde{F}_n(\tau)\log(\tau_1 T_1^2)-\frac{2h_n}{(n+1)^2}\tau^n\tilde{F}_n(\tau)\right\}\\
& +k\sum_{p=1}^{\infty}h_p(\tau_1 T_1^2)^p\sum_{n=1}^{p}\frac{p-n-1}{n}\left\{B_p\left(\frac{\zeta}{T}\right)^{p-n}-D_p\left(\frac{\bar{\zeta}}{T}\right)^{p-n}\right\}\tilde{F}_{p-n}(\tau)\\
& +\sum_{n=0}^{\infty}\frac{s_n}{\frac{1}{2}-n}\left[(n+\tfrac{1}{2})\left(\frac{\zeta}{T}\right)^{\frac{1}{2}-n}\tilde{F}_{\frac{1}{2}-n}(\tau)+\sum_{p=1}^{\infty}(\tfrac{1}{2}+n-p)h_p(\tau_1 T_1^2)^p\right.\\
& \left.\times\left\{B_p\left(\frac{\zeta}{T}\right)^{\frac{1}{2}-n+p}-D_p\left(\frac{\bar{\zeta}}{T}\right)^{\frac{1}{2}-n+p}\right\}\tilde{F}_{\frac{1}{2}-n+p}(\tau)\right]\\
& +k\sum_{n=1}^{\infty}(n-1)h_n(\tau_1 T_1^2)^n\left(\frac{\bar{\zeta}}{T}\right)^n\tilde{F}_n(\tau)\left\{\frac{1}{n^2}+\sum_{p=1}^{\infty}\frac{B_p h_p(\tau_1 T_1^2)^p}{(p+n)^2}\right\}\\
& +k\sum_{n=1}^{\infty}(n-1)h_n(\tau_1 T_1^2)^n\left(\frac{\zeta}{T}\right)^n\tilde{F}_n(\tau)\sum_{p=1}^{\infty}\frac{D_p h_p(\tau_1 T_1^2)^p}{(p+n)^2}. \quad \text{(IV)}
\end{aligned}\tag{7·13}$$

The formula for $i\psi$ in IV′ is derived from this by omitting the term $2\pi ki$ and reversing the sign before the series of terms of half-integral order.

Since the physical solution is

$$\Omega_{\text{phys.}}=iA(\Omega-\bar{\Omega})\quad (A\text{ real}),$$

and in the preceding formulae ψ is the stream function which corresponds to Ω, we have

$$\psi_{\text{phys.}}=iA(\psi-\bar{\psi})=A(i\psi-i\bar{\psi}).$$

Thus (7·1) gives

$$\psi_{\text{phys.}}=A(\psi+\bar{\psi}_0)+A\sum_{p=1}^{\infty}h_p(\tau_1 T_1^2)^p(B_p-D_p)(\psi_p+\bar{\psi}_p),$$

and $\psi_{\text{phys.}}$ is given by adding series, such as those on the right of (7·2) and (7·13), to their conjugates. In particular the term $2\pi ki$ in (7·13) is absent from $\psi_{\text{phys.}}$, which is accordingly single-valued in the plane of the flow.

8. EXPANSIONS OF THE GENERATING FUNCTION AND ITS ASSOCIATES

We start from (3·5):

$$g(\zeta)=-2\sqrt{(\zeta-k'^2)}-2k\log\{k-\sqrt{(\zeta-k'^2)}\},\tag{8·1}$$

which gives

$$g'(\zeta)=\frac{k+\sqrt{(\zeta-k'^2)}}{1-\zeta}.\tag{8·2}$$

On the central line of the domain I we have $k'^2 < \zeta < 1$, and $\surd(\zeta - k'^2)$ is real positive; so, writing

$$(1-z)^{\frac{1}{2}} = \sum_{n=0}^{\infty} c_n z^n,$$

we have in I

$$\begin{aligned} g'(\zeta) &= \left\{k + \zeta^{\frac{1}{2}} \sum_{m=0}^{\infty} c_m (k'^2/\zeta)^m\right\} \sum_{n=0}^{\infty} \zeta^n \\ &= k \sum_{n=0}^{\infty} \zeta^n + \sum_{n=0}^{\infty} \zeta^{\frac{1}{2}+n} \sum_{m=0}^{\infty} c_m k'^{2m} + \sum_{n=0}^{\infty} \zeta^{-\frac{1}{2}-n} \sum_{m=n+1}^{\infty} c_m k'^{2m} \\ &= k \sum_{n=0}^{\infty} \zeta^n + k \sum_{n=0}^{\infty} \zeta^{\frac{1}{2}+n} + \sum_{n=0}^{\infty} \zeta^{-\frac{1}{2}-n} \sum_{m=n+1}^{\infty} c_m k'^{2m}, \end{aligned}$$

since

$$\sum_{0}^{\infty} c_m k'^{2m} = (1-k'^2)^{\frac{1}{2}} = k.$$

Hence

$$g(\zeta) = k \sum_{1}^{\infty} \frac{\zeta^n}{n} + k \sum_{1}^{\infty} \frac{\zeta^{\frac{1}{2}+n}}{\frac{1}{2}+n} + \sum_{0}^{\infty} \frac{q_n \zeta^{\frac{1}{2}-n}}{\frac{1}{2}-n} + C_1, \tag{8·3}$$

where $q_n = \sum_{m=n+1}^{\infty} c_m k'^{2m}$, and the constant C_1 is determined by identifying (8·1) and (8·3) for $\zeta \to k'^2$. Since (8·1) gives $g(k'^2) = -2k \log k$, and the first term on the right of (8·3) becomes $-k \log(1-k'^2) = -2k \log k$, we have

$$\begin{aligned} -C_1 &= k \sum_{1}^{\infty} \frac{k'^{1+2n}}{\frac{1}{2}+n} + \sum_{0}^{\infty} \frac{q_n k'^{1-2n}}{\frac{1}{2}-n} \\ &= k'^3 \sum_{0}^{\infty} c_m k'^{2m} \sum_{0}^{\infty} \frac{k'^{2n}}{\frac{3}{2}+n} + k'^3 \sum_{0}^{\infty} \frac{1}{\frac{1}{2}-n} (c_{n+1} + c_{n+2} k'^2 + c_{n+3} k'^4 + \ldots). \end{aligned} \tag{8·4}$$

For the double series in the last term, the terms for which $n \geqslant 1$ are all positive, and this set of terms is dominated by

$$\sum_{1}^{\infty} \frac{1}{\frac{1}{2}-n} (c_{n+1} + c_{n+2} + \ldots) = \sum_{1}^{\infty} \frac{1}{n-\frac{1}{2}} \frac{1.3.\ldots.(2n-1)}{2.4.\ldots.(2n)(2n+2)},$$

which converges. Hence the whole can be arranged in powers of k'^2, and the coefficient of k'^{3+2r} is

$$\begin{aligned} &\left(\frac{c_0}{\frac{3}{2}+r} + \frac{c_1}{\frac{1}{2}+r} + \ldots + \frac{c_r}{\frac{3}{2}}\right) + \left(\frac{c_{r+1}}{\frac{1}{2}} + \frac{c_{r+2}}{-\frac{1}{2}} + \ldots\right) \\ &= \frac{1}{\frac{3}{2}+r} + \frac{1}{\frac{1}{2}+r} \frac{(-\frac{1}{2})}{1!} + \frac{1}{-\frac{1}{2}+r} \frac{(-\frac{1}{2})(\frac{1}{2})}{2!} + \ldots \\ &= \frac{1}{\frac{3}{2}+r} F(-\tfrac{1}{2}, -\tfrac{3}{2}-r, -\tfrac{1}{2}-r, 1) \\ &= \frac{1}{\frac{3}{2}+r} \frac{\Gamma(-\frac{1}{2}-r)\,\Gamma(\frac{3}{2})}{\Gamma(-r)\,\Gamma(1)} = 0 \quad \text{for} \quad r = 0, 1, 2, \ldots. \end{aligned}$$

Hence $C_1 = 0$. This proves (3·9, I).

To continue $g(\zeta)$ into II we must take, as in (3·6),

$$\surd(\zeta - k'^2) = ik'(1 - \zeta/k'^2)^{\frac{1}{2}}, \tag{8·5}$$

and (8·2) gives
$$g'(\zeta) = \left\{k + ik' \sum_0^\infty c_m(\zeta/k'^2)^m\right\} \sum_0^\infty \zeta^n$$
$$= \sum_0^\infty (k + ik' r_{n+1})\, \zeta^n,$$

where
$$r_n = \sum_{m=0}^{n-1} c_m k'^{-2m}.$$

Hence
$$g(\zeta) = \sum_1^\infty \frac{k + ik' r_n}{n} \zeta^n + C_2, \tag{8·6}$$

where C_2 is most easily found by identifying (8·1) and (8·6) at $\zeta = 0$:

$$C_2 = g(0) = -2ik' - 2k \log(k - ik') \quad \text{(from (8·5))},$$
$$= -2ik' + 2ik \arcsin k' = -\tfrac{2}{3} ik'^3 F(1, 1, \tfrac{5}{2}, k'^2). \tag{8·7}$$

The form last given may be proved, for example, by forming a linear differential equation for $(1 - k'^2)^{\frac{1}{2}} \arcsin k'$ and solving it in series. This proves (3·9, II), and for III we have merely to reverse the sign before i.

For the expansion of $g'(\zeta)$ in I′ we reverse the signs of the terms of half-integral order in the expansion for I, so that in I′

$$g(\zeta) = k \sum_1^\infty \frac{\zeta^n}{n} - k \sum_1^\infty \frac{\zeta^{\frac{1}{2}+n}}{\frac{1}{2}+n} - \sum_0^\infty \frac{q_n \zeta^{\frac{1}{2}-n}}{\frac{1}{2}-n} + C_3. \tag{8·8}$$

The constant C_3 is found by putting $\zeta = k'^2$. Then, as already shown,

$$g(k'^2) = k \sum_1^\infty \frac{k'^{2n}}{n}, \quad k \sum_1^\infty \frac{k'^{1+2n}}{\frac{1}{2}+n} + \sum_0^\infty \frac{q_n k'^{1-2n}}{\frac{1}{2}-n} = 0,$$

so $C_3 = 0$. This proves (3·9, I′).

For IV, which is to be entered from I, we observe that in (8·3) the last series converges for $|\zeta| > k'^2$, so we have only to find the continuation of the first two series, viz.

$$k\left(\sum_1^\infty \frac{\zeta^n}{n} + \sum_1^\infty \frac{\zeta^{\frac{1}{2}+n}}{\frac{1}{2}+n}\right) = 2k \sum_2^\infty \frac{(\zeta^{\frac{1}{2}})^n}{n} = -2k\{\log(1 - \zeta^{\frac{1}{2}}) + \zeta^{\frac{1}{2}}\}, \tag{8·9}$$

where $\zeta^{\frac{1}{2}}$ is the principal value. If ζ circles in the positive sense round $\zeta = 1$, then $\zeta^{\frac{1}{2}}$ circles in the positive sense round $\zeta^{\frac{1}{2}} = 1$, and in IV we are to put

$$\log(1 - \zeta^{\frac{1}{2}}) = \pi i + \log(\zeta^{\frac{1}{2}} - 1)$$
$$= \pi i + \tfrac{1}{2}\log\zeta + \log(1 - \zeta^{-\frac{1}{2}})$$
$$= \pi i + \tfrac{1}{2}\log\zeta - \zeta^{-\frac{1}{2}} - \tfrac{1}{2}\zeta^{-1} - \tfrac{1}{3}\zeta^{-\frac{3}{2}} - \ldots.$$

Hence the continuation of (8·9) is

$$-k(2\pi i + \log\zeta) + k\left(\zeta + \frac{\zeta^2}{2} + \frac{\zeta^3}{3} + \ldots\right) - k\left(\frac{\zeta^{\frac{1}{2}}}{\frac{1}{2}} - \frac{\zeta^{-\frac{1}{2}}}{\frac{1}{2}} - \frac{\zeta^{-\frac{3}{2}}}{\frac{3}{2}} - \ldots\right)$$
$$= -k(2\pi i + \log\zeta) + k \sum_1^\infty \frac{\zeta^{-n}}{n} - k \sum_0^\infty \frac{\zeta^{\frac{1}{2}-n}}{\frac{1}{2}-n}.$$

On adding the last series in (8·3) we obtain in IV

$$g(\zeta) = -k(2\pi i + \log\zeta) + k\sum_1^\infty \frac{\zeta^{-n}}{n} - \sum_0^\infty \frac{\zeta^{\frac{1}{2}-n}}{\frac{1}{2}-n}\left(\sum_0^\infty c_m k'^{2m} - \sum_{n+1}^\infty c_m k'^{2m}\right)$$

$$= -k(2\pi i + \log\zeta) + k\sum_1^\infty \frac{\zeta^{-n}}{n} - \sum_0^\infty \frac{s_n \zeta^{\frac{1}{2}-n}}{\frac{1}{2}-n}, \tag{8·10}$$

where $$s_n = \sum_0^n c_m k'^{2m}.$$

This proves (3·9, IV).

For IV′ the expansion is to be obtained by continuation from I′, so in the preceding investigation for IV we have to reverse the sign of $\zeta^{\frac{1}{2}}$. In place of

$$\log(1-\zeta^{\frac{1}{2}}) = \pi i + \log(\zeta^{\frac{1}{2}} - 1) = \pi i + \tfrac{1}{2}\log\zeta + \log(1-\zeta^{-\frac{1}{2}}),$$

we now have $$\log(1+\zeta^{\frac{1}{2}}) = \log(\zeta^{\frac{1}{2}}+1) = \tfrac{1}{2}\log\zeta + \log(1+\zeta^{-\frac{1}{2}});$$

so we get the expansion for IV′ from that for I′ by reversing the sign of $\zeta^{\frac{1}{2}}$ and also omitting the term $-2\pi i k$. This proves (3·9, IV′).

The associated functions

The function $g_m^{(p)}(\zeta)$ is to be regarded as defined in the domain I by the expansion (3·18), viz.

$$g_m^{(p)}(\zeta) = k\sum_{n=1}^\infty \frac{\zeta^{n+p}}{n(n+p+m)} + k\sum_1^\infty \frac{\zeta^{\frac{1}{2}+n+p}}{(\frac{1}{2}+n)(\frac{1}{2}+n+p+m)} + \sum_0^\infty \frac{q_n \zeta^{\frac{1}{2}-n+p}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)}; \tag{8·11}$$

and it is to be thence determined in II, III, I′ by continuation within $|\zeta| < 1$, in IV by continuation from I, and in IV′ by continuation from I′. Since (8·11) gives

$$\frac{d}{d\zeta}\zeta^m g_m^{(p)}(\zeta) = \zeta^{p+m-1}\left\{k\sum_1^\infty \frac{\zeta^n}{n} + k\sum_1^\infty \frac{\zeta^{\frac{1}{2}+n}}{\frac{1}{2}+n} + \sum_0^\infty \frac{q_n\zeta^{\frac{1}{2}-n}}{\frac{1}{2}-n}\right\},$$

$$= \zeta^{p+m-1} g(\zeta) \quad \text{by (8·3)},$$

the continuations already found for $g(\zeta)$ furnish those for $(d/d\zeta)(\zeta^m g_m^{(p)}(\zeta))$, and the various series for $\zeta^m g_m^{(p)}(\zeta)$ are thence found, to an additive constant.

Thus in II we have, from (8·6),

$$g_m^{(p)}(\zeta) = \sum_{n=1}^\infty \frac{(k + ik' r_n)\zeta^{p+n}}{n(p+m+n)} + C_2\frac{\zeta^p}{p+m} + C_4\zeta^{-m}, \tag{8·12}$$

where C_4 is to be found by identifying this series with (8·11) for $\zeta = k'^2$. Hence

$$C_4 k'^{-2m} + C_2\frac{k'^{2p}}{p+m} + ik'\sum_1^\infty \frac{r_n k'^{2p+2n}}{n(p+m+n)}$$

$$= k\sum_1^\infty \frac{k'^{1+2n+2p}}{(\frac{1}{2}+n)(\frac{1}{2}+n+p+m)} + \sum_0^\infty \frac{q_n k'^{1-2n+2p}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)}. \tag{8·13}$$

The series on the right is

$$\frac{k'^{2p}}{p+m}\left\{k\sum_1^\infty \frac{k'^{1+2n}}{\frac{1}{2}+n} + \sum_0^\infty \frac{q_n k'^{1-2n}}{\frac{1}{2}-n}\right\} - \frac{k'^{2p}}{p+m}\left\{k\sum_1^\infty \frac{k'^{1+2n}}{\frac{1}{2}+n+p+m} + \sum_0^\infty \frac{q_n k'^{1-2n}}{\frac{1}{2}-n+p+m}\right\}.$$

The series in the first $\{\ldots\}$ is the value of $-C_1$ in (8·4), which has been proved to be zero. For the series in the other $\{\ldots\}$ the coefficient of k'^{3+2r} reduces similarly to

$$\frac{c_0}{\frac{3}{2}+p+m+r}+\frac{c_1}{\frac{1}{2}+p+m+r}+\ldots=\frac{1}{\frac{3}{2}+p+m+r}F(-\tfrac{1}{2},-\tfrac{3}{2}-p-m-r,-\tfrac{1}{2}-p-m-r,1),$$

which is 0, since p, m are positive integers and $r=0,1,2,\ldots$.

Also, writing $p+m=t$,

$$k'\sum_1^\infty\frac{r_nk'^{2p+2n}}{n(p+m+n)}=k'^{1+2p}\sum_1^\infty\frac{c_0k'^{2n}+c_1k'^{2n-2}+\ldots+c_{n-1}k'^2}{n(t+n)},$$

where the double series converges absolutely since

$$|c_0|+|c_1|+\ldots+|c_{n-1}|=1+|c_1+c_2+\ldots+c_{n-1}|<2.$$

Hence we can arrange it in powers of k'^2, and the coefficient of $k'^{1+2p+2r}$ is

$$\frac{c_0}{r(t+r)}+\frac{c_1}{(r+1)(t+r+1)}+\ldots=\frac{1}{t}\left\{\frac{c_0}{r}+\frac{c_1}{r+1}+\ldots\right\}-\frac{1}{t}\left\{\frac{c_0}{t+r}+\frac{c_1}{t+r+1}+\ldots\right\},$$

where
$$\begin{aligned}\frac{c_0}{t+r}+\frac{c_1}{t+r+1}+\ldots&=\frac{1}{t+r}+\frac{(-\frac{1}{2})}{t+r+1}+\frac{(-\frac{1}{2})(\frac{1}{2})}{(t+r+1)\,2!}+\ldots\\&=\frac{1}{t+r}F(-\tfrac{1}{2},t+r,t+r+1,1)\\&=\frac{1}{t+r}\frac{\Gamma(t+r+1)\,\Gamma(\frac{3}{2})}{\Gamma(t+r+\frac{3}{2})\,\Gamma(1)}=\frac{\Gamma(t+r)\,\Gamma(\frac{3}{2})}{\Gamma(t+r+\frac{3}{2})}.\end{aligned}$$

Since r runs over the values $1,2,\ldots$, we have then

$$\begin{aligned}k'\sum_1^\infty\frac{r_nk'^{2p+2n}}{n(p+m+n)}&=\frac{k'^{1+2p}}{t}\left\{\sum_{n=1}^\infty\frac{\Gamma(n)\,\Gamma(\frac{3}{2})}{\Gamma(n+\frac{3}{2})}k'^{2n}-\sum_{n=1}^\infty\frac{\Gamma(t+n)\,\Gamma(\frac{3}{2})}{\Gamma(t+n+\frac{3}{2})}k'^{2n}\right\}\\&=\frac{k'^{3+2p}}{t}\left\{\frac{\Gamma(\frac{3}{2})}{\Gamma(\frac{5}{2})}F(1,1,\tfrac{5}{2},k'^2)-\frac{\Gamma(\frac{3}{2})\,\Gamma(t+1)}{\Gamma(t+\frac{5}{2})}F(1,t+1,t+\tfrac{5}{2},k'^2)\right\}.\end{aligned}$$

Replacing t by $p+m$, (8·13) gives therefore

$$C_4k'^{-2m}+C_2\frac{k'^{2p}}{p+m}=\frac{ik'^{3+2p}\Gamma(\frac{3}{2})\,\Gamma(p+m)}{\Gamma(p+m+\frac{5}{2})}F(1,1+p+m,\tfrac{5}{2}+p+m,k'^2)-\frac{\frac{2}{3}ik'^{3+2p}}{p+m}F(1,1,\tfrac{5}{2},k'^2),$$

and on account of (8·7) we have

$$C_4=i\frac{k'^{3+2p+2m}\,\Gamma(\frac{3}{2})\,\Gamma(p+m)}{\Gamma(\frac{5}{2}+p+m)}F(1,1+p+m,\tfrac{5}{2}+p+m,k'^2).$$

The expansion (8·12) is thus determined, and (5·2, II) is proved. The expansion for III is the conjugate of this.

For I′, on integrating the appropriate expansion of $\zeta^{p+m-1}g(\zeta)$, we obtain

$$g_m^{(p)}(\zeta)=k\sum_1^\infty\frac{\zeta^{n+p}}{n(n+p+m)}-k\sum_1^\infty\frac{\zeta^{\frac{1}{2}+n+p}}{(\frac{1}{2}+n)(\frac{1}{2}+n+p+m)}-\sum_0^\infty\frac{q_n\zeta^{\frac{1}{2}-n+p}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)}+C_5\zeta^{-m},$$

where the constant C_5 is to be obtained by identifying this with (8·11) for $\zeta=k'^2$. Since we have proved that this substitution annuls the series of half-integral powers, we have $C_5=0$, and (5·2, I′) is proved.

For IV, the appropriate expansion of $\zeta^{p+m-1}g(\zeta)$ is given by (8·10), and on integrating this we get

$$\begin{aligned} g_m^{(p)}(\zeta) = {} & -\frac{k(2\pi i+\log\zeta)\,\zeta^p}{p+m}+\frac{k\zeta^p}{(p+m)^2}+k\sum_1^{\infty}{}'\frac{\zeta^{p-n}}{n(p+m-n)} \\ & +\frac{k\log\zeta}{p+m}\zeta^{-m}-\sum_0^{\infty}\frac{s_n\zeta^{\frac{1}{2}-n+p}}{(\frac{1}{2}-n)(\frac{1}{2}-n+p+m)}+C_6\zeta^{-m}, \end{aligned}$$

and we determine C_6 by identifying this with (8·11) at $\zeta = 1$, both series being continuous up to this point, by Abel's theorem. Putting $p+m = t$, this gives

$$C_6 = \frac{2\pi ik}{t}-\frac{k}{t^2}-k\sum_1^{\infty}{}'\frac{1}{n(t-n)}+k\sum_1^{\infty}\frac{1}{n(t+n)}+k\sum_1^{\infty}\frac{1}{(\frac{1}{2}+n)(\frac{1}{2}+t+n)}+\sum_0^{\infty}\frac{q_n+s_n,}{(\frac{1}{2}-n)(\frac{1}{2}+t-n)}, \tag{8·14}$$

and here $$q_n+s_n = \sum_{n+1}^{\infty}c_m k'^{2m}+\sum_0^{n}c_m k'^{2m} = (1-k'^2)^{\frac{1}{2}} = k.$$

The last two terms therefore combine into

$$k\sum_{-\infty}^{\infty}\frac{1}{(\frac{1}{2}+n)(\frac{1}{2}+n+t)} = \frac{k}{t}\sum_{-\infty}^{\infty}\left(\frac{1}{\frac{1}{2}+n}-\frac{1}{\frac{1}{2}+n+t}\right) = 0,$$

since t is an integer. The preceding two terms combine into

$$k\sum_{-\infty}^{\infty}{}''\frac{1}{n(n+t)} = \frac{k}{t}\sum_{-\infty}^{\infty}{}''\left(\frac{1}{n}-\frac{1}{n+t}\right),$$

where Σ'' omits the terms with $n = 0$ and $n = -t$; and this is equal to

$$\frac{k}{t}\frac{2}{t}.$$

Hence $$C_6 = \frac{2\pi ik}{t}+\frac{k}{t^2} = \frac{2\pi ik}{p+m}+\frac{k}{(p+m)^2},$$

and (5·2, IV) is proved. The expansion (5·2, IV′) is proved similarly.

REFERENCES

Cherry, T. M. 1947 *Proc. Roy. Soc.* A, **192**, 45.
Lighthill, M. J. 1947 *Proc. Roy. Soc.* A, **191**, 352.

Numerical solutions for transonic flow

By T. M. Cherry, *The University of Melbourne*

(*Communicated by G. Temple, F.R.S.—Received* 15 *July* 1948)

In a previous paper (Cherry 1947), the author has established a family of exact solutions for steady two-dimensional flow of a compressible fluid past a cylinder; the final formulae are given in theorem 6, equations (5·17) to (5·21). These formulae have now been evaluated (taking $\gamma = 1{\cdot}405$) for the value $\tau_1 = 0{\cdot}05$, corresponding to a free-stream Mach number of 0·510, and the streamlines are shown in figure 1. The cylindrical obstacle has a thickness ratio 0·93, but is markedly different from

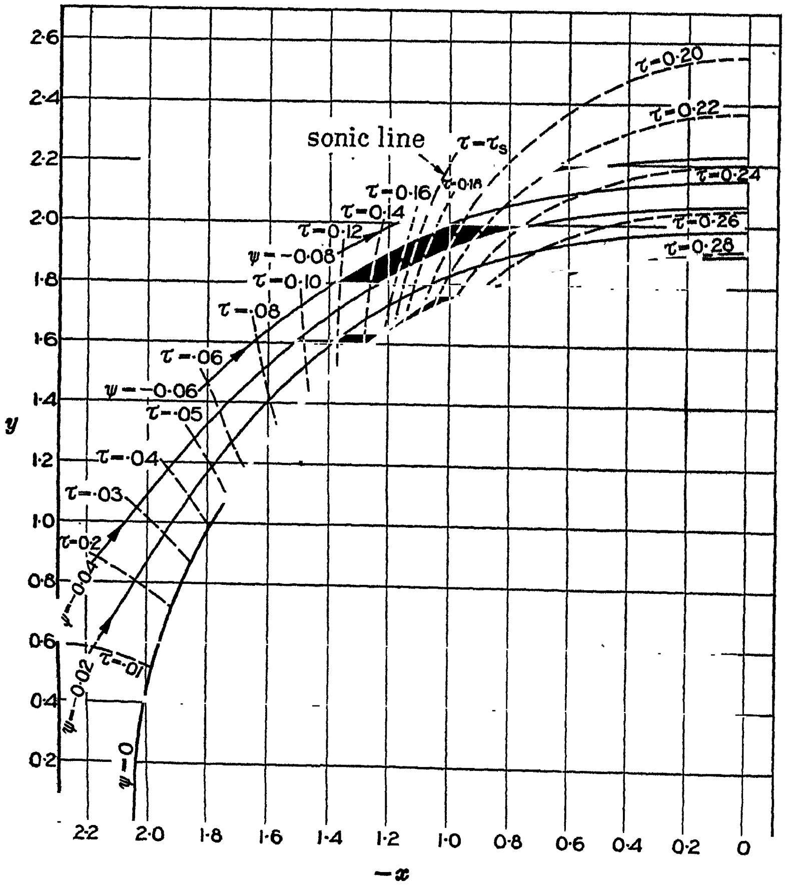


Figure 1. —— Streamlines and - - - lines of constant speed, density and pressure for solution P. The left-hand part of the streamline $\psi = 0$ is very close to a circle centred at $x = 0$, $y = 0$; at about $x = -1{\cdot}3$, $y = 1{\cdot}6$ the circle is shown diverging from the streamline.

an ellipse, being almost exactly circular over its up- and downstream quadrants. The Mach number at the ends of its transverse axis is 1·39. The flow is everywhere regular, but a small increase in the free-stream Mach number would be critical; a shock-line would begin to appear near the points on the surface where the tangent is inclined at about 25 or 30° to the direction of the free-stream.

Since the hodograph equations are linear, new solutions can be obtained by superposing two or more known solutions. In this way the solutions illustrated in figures 2 and 3 have been obtained. Calling P the solution of figure 1, figure 2 shows the solution $P+0{\cdot}94C_2$, and figure 3 the solution $P+1{\cdot}50S_2$, where C_2, S_2 are the cases $\nu=2$ of the following:

$$C_\nu \equiv \begin{cases} \psi = -\dfrac{\nu-1}{\nu}\tau^{\frac{1}{2}\nu}\tilde{F}_\nu(\tau)\cos\nu\theta, \\[2ex] x = \tau^{\frac{1}{2}\nu-\frac{1}{2}}\left(\dfrac{2\tau F'_\nu(\tau)}{\nu}\cos\theta\,\sin\nu\theta + F_\nu(\tau)\sin(\nu-1)\,\theta\right), \\[2ex] y = \tau^{\frac{1}{2}\nu-\frac{1}{2}}\left(\dfrac{2\tau F'_\nu(\tau)}{\nu}\sin\theta\,\sin\nu\theta + F_\nu(\tau)\cos(\nu-1)\,\theta\right), \end{cases}$$

$$S_\nu \equiv \begin{cases} \psi = \dfrac{\nu-1}{\nu}\tau^{\frac{1}{2}\nu}\tilde{F}_\nu(\tau)\sin\nu\theta, \\[2ex] x = \tau^{\frac{1}{2}\nu-\frac{1}{2}}\left(\dfrac{2\tau F'_\nu(\tau)}{\nu}\cos\theta\,\cos\nu\theta + F_\nu(\tau)\cos(\nu-1)\,\theta\right), \\[2ex] y = \tau^{\frac{1}{2}\nu-\frac{1}{2}}\left(\dfrac{2\tau F'_\nu(\tau)}{\nu}\sin\theta\,\cos\nu\theta - F_\nu(\tau)\sin(\nu-1)\,\theta\right). \end{cases}$$

The constants 0·94, 1·50 were so chosen that these new solutions should be critical; in each case there are nascent shock-lines *on* the surface of the cylinder, and the continuation of the flow within the surface shows limit-lines. At these critical points the curvature of the surface is infinite, but the slope is continuous.

For $P+0{\cdot}94C_2$ the greatest Mach number in the field is 1·56, and at the critical point it is 1·335. For $P+1{\cdot}50S_2$ the greatest Mach number in the field is 1·60, and at the critical point it is 1·38. In both cases the acceleration (which becomes infinite at the critical point and then changes sign) is markedly unsymmetrical about the critical point.

That the streamline through a critical point should have its curvature there infinite is not, of course, a general physical result; it is simply a feature of those critical solutions for transonic flow which are *most easily obtained* by the hodograph method, and this is the only known exact method.

The numerical work underlying figures 1 to 3 is heavy. The main feature is that slowly convergent series have to be evaluated, and this requires the estimation of a 'remainder' after summing the early terms. For the final results there is an independent check, which indicates that they are accurate to three significant figures. The details of the procedure, and the numerical results, are given in Report A48 (by T. M. Cherry) of the C.S.I.R. Division of Aeronautics, Fishermen's Bend,

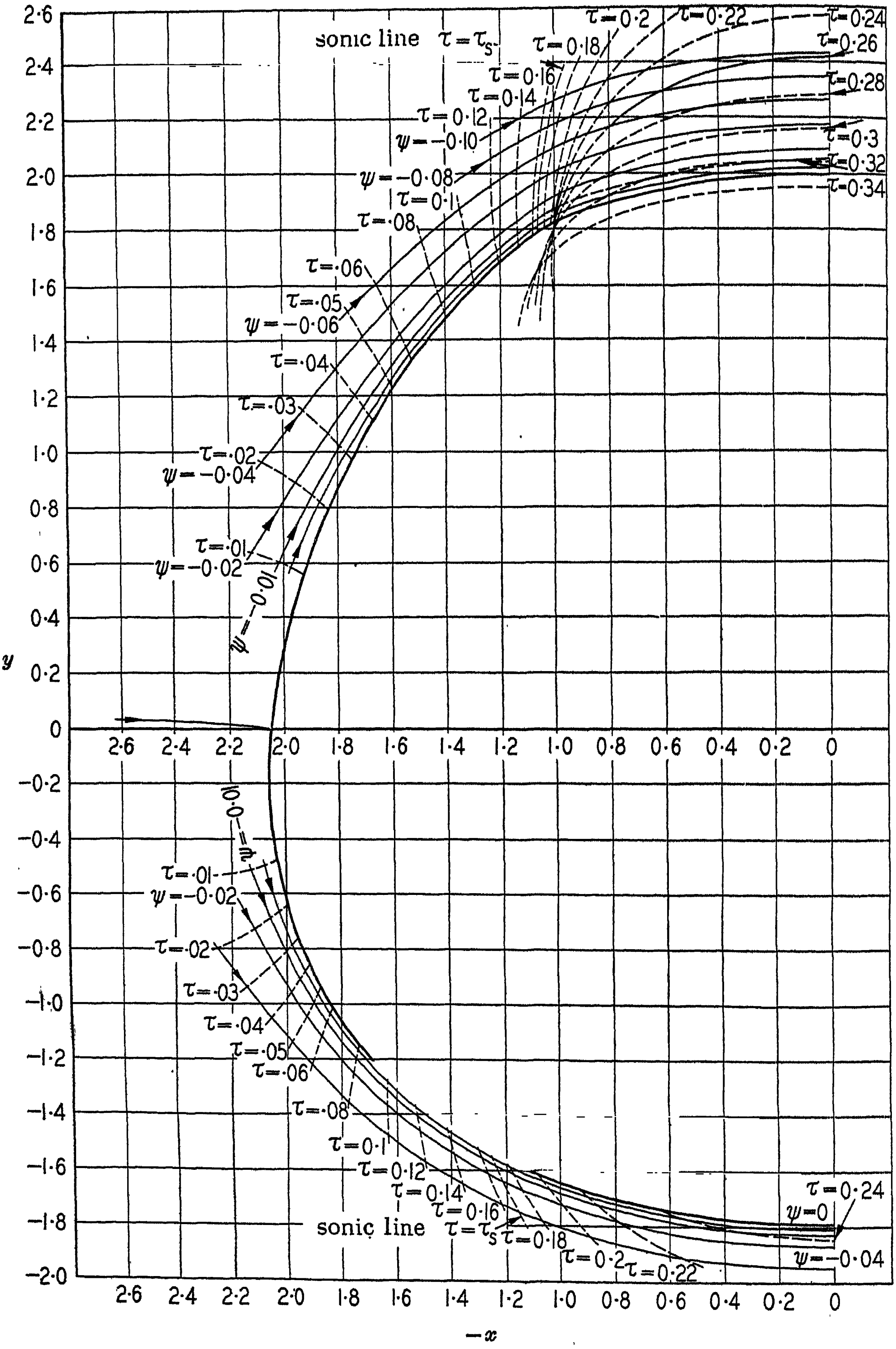


Figure 2. —— Streamlines and - - - lines of constant speed, density and pressure for solution $P+0{\cdot}94C_2$.

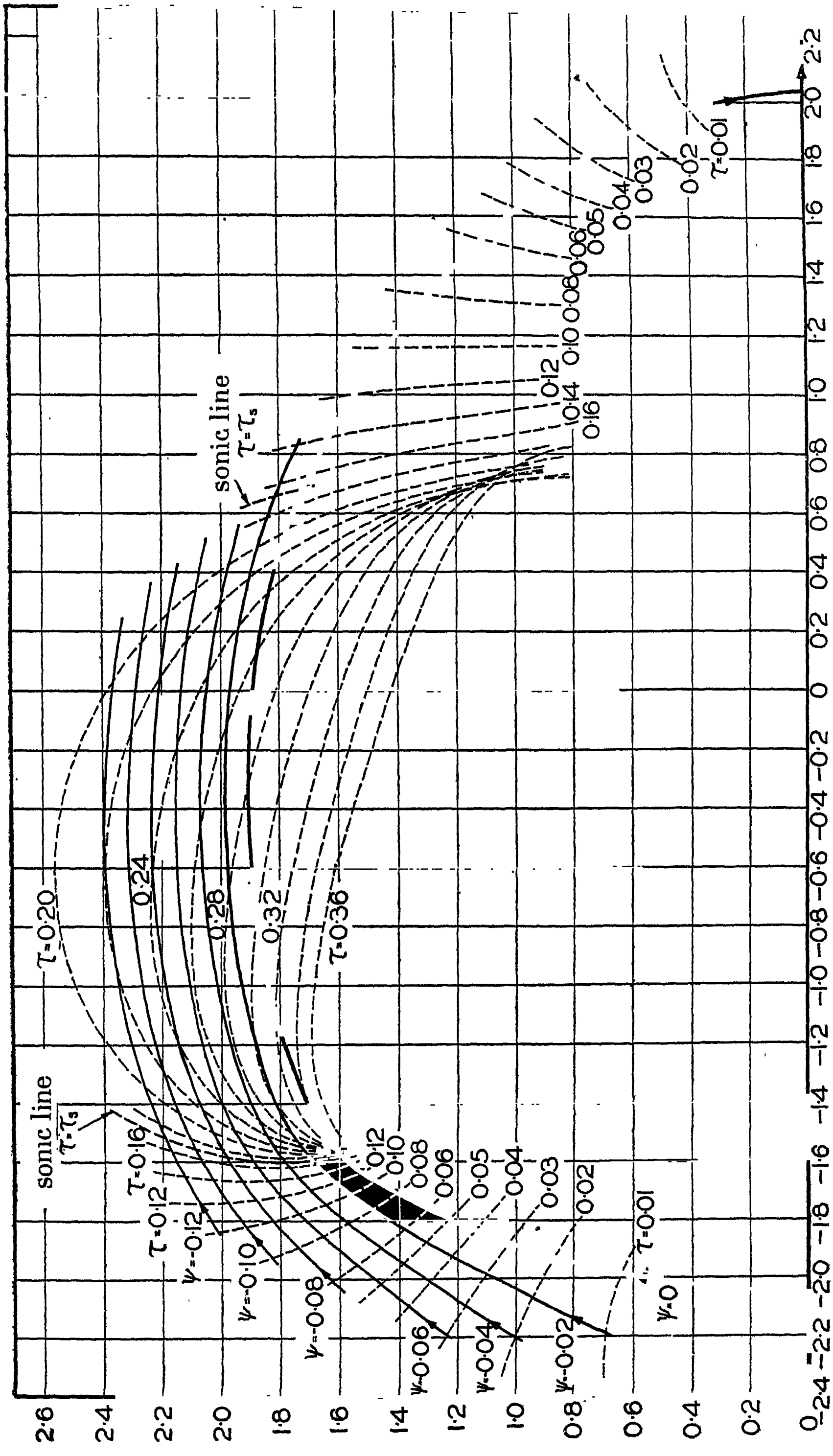


FIGURE 3. —— Streamlines and - - - lines of constant speed, density and pressure for solution $P+1{\cdot}50S_2$.

Melbourne, Australia. The calculations were confined to the most interesting part of the field, which explains the incompleteness of the outer streamlines in the figures.

Acknowledgement is gratefully made to Miss Betty Laby and Mr H. C. Levey for their help in this project, and to the Division of Aeronautics for making their services available.

REFERENCE

Cherry, T. M. 1947 *Proc. Roy. Soc.* A, 192, 45.

Order-disorder statistics. I

BY C. DOMB, *Pembroke College, University of Cambridge*

(*Communicated by D. R. Hartree, F.R.S.—Received* 8 *May* 1948—
Revised 24 *August* 1948)

In 1941 Kramers & Wannier discussed the statistical mechanics of a two-dimensional Ising model of a ferromagnetic. By making use of a 'screw transformation' they showed that the partition function was the largest eigenvalue of an infinite matrix of simple characteristic structure. In the present paper an alternative method is used for deriving the partition function, and this enables the 'screw transformation' to be generalized to apply to a number of problems of classical statistical mechanics, including the three-dimensional Ising model. Distant neighbour interactions can also be taken into account. The relation between the ferromagnetic and order-disorder problems is discussed, and it is shown that the partition function in both cases can be derived from a single function of two variables. Since distant neighbour interactions can be taken into account the theory can be formally applied to the statistical mechanics of a system of identical particles.

1. INTRODUCTION

There are a large number of problems in statistical mechanics in which a classical solution furnishes a satisfactory approximation. When this is the case, if our assembly consists of a large number of particles, the partition function for the assembly reduces to a product of two factors, a kinetic energy factor, and a configurational factor. The first of these can easily be evaluated, and the fundamental problem which remains is the determination of the configurational partition function. Examples of problems of this type are the condensation and solidification of many substances, the order-disorder transitions in alloys, the theory of regular solid solutions, and the behaviour of ferromagnetic substances on the basis of an Ising model. In the present paper, we shall be largely concerned with the last three phenomena although § 5 will have some bearing on the first of them.

In the order-disorder* problem calculations are usually made on the assumption that the individual units are arranged in a given lattice structure. The potential energy arises from the short-ranged interatomic forces, and it is quite reasonable to take as a first working model one in which only nearest neighbours interact. The energy of a given configuration is completely determined by specifying the number of nearest neighbour pairs of all possible types, and the mathematical problem to be solved is the determination of the number of configurations involving a specified number of nearest neighbour pairs of each type. For example, if we have an alloy composed of n_1 atoms of type 1 and n_2 atoms of type 2, there are three possible nearest neighbour types, 1–1, 1–2 and 2–2, and our problem reduces to determining the number $h(n_1, n_2; n_{11}, n_{12}, n_{22})$ of configurations involving n_{11} pairs of the first type, n_{12} of the second, and n_{22} of the third. Similar considerations apply to the Ising model of a ferromagnetic. Each spin in the lattice is capable of two orientations which may occur with equal probability, and our mathematical problem is again one of determining the average number of configurations containing specified numbers of each orientation, and of nearest neighbour types. Except in the case of a linear chain the mathematical problem described above is one of some difficulty, and various approximations have been devised; as a typical example we may quote the Bethe approximation in the order-disorder problem (Bethe 1935). These approximations predict a discontinuity in specific heat at a given temperature (the Curie point), and were at first considered fairly satisfactory. The first attempt at obtaining exact information as to the position and nature of the singularity was made by Kramers & Wannier (1941). In considering the problem of a two-dimensional ferromagnetic in the absence of a magnetic field, they showed that the partition function was the largest eigenvalue of an infinite matrix of characteristic structure (a V-matrix). By using symmetry properties of this matrix they located the Curie point, and by means of finite matrix approximations they conjectured that the specific heat had a logarithmic infinity at this point. These results were completely confirmed by Onsager (1944), when, using group theoretical methods, he obtained a complete solution to this problem. This meant, however, that the various approximations which had been previously employed were appreciably wrong both as regards the location and nature of the Curie point. Many of these approximations, and their relation to the exact solution, are reviewed in an article by Wannier (1945).

The approach employed in the present paper is similar to that of Kramers & Wannier. It is shown that a number of problems can be reduced to the determination of the largest eigenvalue of an infinite matrix of characteristic structure (a duodiagonal matrix). These include the three-dimensional ferromagnetic in the presence of a magnetic field, and the general order-disorder problem with any given ratio of the two constituents. Distant neighbour interactions can be taken into account and do not alter the characteristic structure of the matrix. It seems, therefore, that a study of matrices of this structure should provide useful information regarding the nature of the solution of problems of the above type, and it is hoped to undertake this subsequently.

---

* In the present paper we shall not be concerned with the detailed calculation of physical results, but with the transformation of a number of physical problems into standard mathematical form. Thus the term 'order-disorder' includes also solid solutions, and the term 'ferromagnetic' includes also anti-ferromagnetic. These problems differ in the sign of the interaction energy.

The relation between the order-disorder and ferromagnetic problems is discussed in § 2, and it is shown that both solutions can be derived from a single function of two variables.

2. THE LINEAR CHAIN

It is well known that a one-dimensional chain does not exhibit any phase transitions similar to those characteristic of two- and three-dimensional structures. Nevertheless, since an exact solution is quite easy in this case, it serves as a useful illustration of the general method to be used. Let us first consider a problem in probability. Suppose we are given n units in a line, each unit being capable of two configurations, 1 and 2, the probabilities of these configurations being α_1 and α_2 respectively ($\alpha_1+\alpha_2 = 1$); what is the probability of a total configuration involving n_1 units of type 1, n_2 units of type 2, and n_{11}, n_{12}, n_{22} nearest neighbour pairs of types 1–1, 1–2, 2–2 respectively? When $\alpha_1 = \alpha_2 = \frac{1}{2}$, this is essentially the one-dimensional ferromagnetic problem.

For large n we may take n_1, n_{12} as the two independent numbers of the above five and the remaining three can be deduced from them by the relations $n_{11}+\frac{1}{2}n_{12} = n_1$, $\frac{1}{2}n_{12}+n_{22} = n_2$, $n_1+n_2 = n$. Let $w(n;\, n_1, n_{12})$ be the probability of a configuration of the required type. We can divide these configurations into two mutually exclusive classes, those involving 1 in the last unit, probability $w_1(n;\, n_1, n_{12})$, and those involving 2 in the last unit, probability $w_2(n;\, n_1, n_{12})$. Then

$$w(n;\, n_1, n_{12}) = w_1(n;\, n_1, n_{12}) + w_2(n;\, n_1, n_{12}).$$

It is now easy to deduce the following equations

$$\left.\begin{aligned} w_1(n+1;\, n_1, n_{12}) &= \alpha_1 w_1(n;\, n_1-1, n_{12}) + \alpha_1 w_2(n;\, n_1-1, n_{12}-1),\\ w_2(n+1;\, n_1, n_{12}) &= \alpha_2 w_1(n;\, n_1, n_{12}-1) + \alpha_2 w_2(n;\, n_1, n_{12}).\end{aligned}\right\} \qquad (1)$$

In statistical mechanics it is more usual to deal with the average number of configurations of a specified type rather than the probability of these configurations. Denoting this by $g(n;\, n_1, n_{12})$ since the total number of configurations is 2^n, we have $g(n;\, n_1, n_{12}) = 2^n w(n;\, n_1, n_{12})$, so that

$$\left.\begin{aligned} g_1(n+1;\, n_1, n_{12}) &= 2\alpha_1 g_1(n;\, n_1-1, n_{12}) + 2\alpha_1 g_2(n;\, n_1-1, n_{12}-1),\\ g_2(n+1;\, n_1, n_{12}) &= 2\alpha_2 g_1(n;\, n_1, n_{12}-1) + 2\alpha_2 g_2(n;\, n_1, n_{12}).\end{aligned}\right\} \qquad (2)$$

The usual method of solving such a set of linear difference equations is to consider the functions

$$f_1(x;\, x_1, x_{12}) = \sum_1^\infty g_1(n;\, n_1, n_{12})\, x_1^{n_1} x_{12}^{n_{12}} x^n,$$

and

$$f_2(x;\, x_1, x_{12}) = \sum_1^\infty g_2(n;\, n_1, n_{12})\, x_1^{n_1} x_{12}^{n_{12}} x^n.$$

It is then easy to show that $g(n; n_1, n_{12})$ is the coefficient of $x_1^{n_1} x_{12}^{n_{12}} x^n$ in the expansion of

$$f(x; x_1, x_{12}) = f_1(x; x_1, x_{12}) + f_2(x; x_1, x_{12}) = \frac{\phi(x; x_1, x_{12})}{(2\alpha_1 x_1 x - 1)(2\alpha_2 x - 1) - x_1 x_{12}^2 x^2}, \tag{3}$$

where $\phi(x; x_1, x_{12})$ is a polynomial of small order determined by the first terms. We can deal with the term in x by separating (3) into partial fractions in x and expanding in powers of x in the standard manner. We thus deduce that $g(n; n_1, n_{12})$ is the coefficient of $x_1^{n_1} x_{12}^{n_{12}}$ in

$$A_1 \lambda_1^n + A_2 \lambda_2^n, \tag{4}$$

where λ_1, λ_2 are the roots of

$$\begin{vmatrix} 2\alpha_1 x_1 - \lambda & 2\alpha_1 x_1 x_{12} \\ 2\alpha_2 x_{12} & 2\alpha_2 - \lambda \end{vmatrix} = 0, \tag{5}$$

and A_1, A_2 are functions of x_1, x_{12}.

The partition of function for n units is defined as

$$Z = \sum_{n_1, n_{12}} g(n; n_1, n_{12}) \exp\left[-\epsilon(n; n_1, n_{12})/kT\right],$$

where $\epsilon(n; n_1, n_{12})$ is the energy of the $(n; n_1, n_{12})$ configuration. We assume that this energy can be expressed as a linear function of n, n_1, n_{12},

$$\epsilon = \epsilon_0 n + \epsilon_1 n_1 + \epsilon_{12} n_{12}. \tag{6}$$

The term in n can be taken out of the summation and effectively neglected, and we see that if we replace x by $\exp(-\epsilon_1/kT)$ and x_{12} by $\exp(-\epsilon_{12}/kT)$ in (4) we obtain the partition function. Also for large n only the largest root, λ_1, need be considered, since the other term is negligible by comparison. Hence we may write for the free energy F,

$$F = -kT \log Z = -nkT \log \lambda_1, \tag{7}$$

where λ_1 is the largest eigenvalue of the matrix

$$\begin{vmatrix} 2\alpha_1 \mu & 2\alpha_1 \mu z \\ 2\alpha_2 z & 2\alpha_2 \end{vmatrix}, \tag{8}$$

and $$\mu = \exp(-\epsilon_1/kT), \quad z = \exp(-\epsilon_{12}/kT).$$

The partition function for a linear ferromagnetic follows immediately from (8) if we put $\alpha_1 = \alpha_2 = \frac{1}{2}$, identify $\frac{1}{2}\epsilon_{12}$ with the energy of interaction of two similarly directed spins, and $\frac{1}{2}\epsilon_1$ with the energy of interaction of a single spin in a magnetic field. This is in agreement with the solution derived previously by other authors (Kramers & Wannier 1941; Montroll 1942).

In the order-disorder problem we do not deal with units capable of both configurations, and with certain probabilities of taking on each, but instead we are told definitely that there are n_1 units of one type and n_2 of the other. One might now be tempted, by analogy with problems of this kind arising in statistics, to identify the solution for $\mu = 1$, $\epsilon_1 = 0$ with the order-disorder problem, the fractions of the two constituents being given by α_1 and α_2. This procedure is permissible in ordinary statistical problems since we are prepared to neglect small probabilities; but in the present problem small probabilities are important at low temperatures, and the step is not justifiable. One can see that the problems are not identical by considering the lowest energy state at zero temperature. In the probability case the lowest state is always non-degenerate, and equivalent to the lowest energy state with equal ratio of constituents; whereas in the order-disorder case with unequal ratio of constituents the lowest energy state is degenerate. We proceed to discuss the order-disorder problem, and to obtain the correct solution.

Let $h(n; n_1, n_{12})$ be the number of configurations involving n_1 units of the first kind, and n_{12} nearest neighbour pairs of type 1–2. We divide $h(n; n_1, n_{12})$ into mutually exclusive groups $h_1(n; n_1, n_{12})$ and $h_2(n; n_1, n_{12})$. The question of probability does not enter here, and in any given case n_1 is fixed. We deduce the equations*

$$\left.\begin{aligned} h_1(n+1; n_1, n_{12}) &= h_1(n; n_1-1, n_{12}) + h_2(n; n_1-1, n_{12}-1), \\ h_2(n+1; n_1, n_{12}) &= h_1(n; n_1, n_{12}-1) + h_2(n; n_1, n_{12}). \end{aligned}\right\} \tag{9}$$

Proceeding in an exactly analogous manner to the previous case equations (4) and (5) are valid for $h(n; n_1, n_{12})$ with $2\alpha_1 = 2\alpha_2 = 1$. The partition function, however, is now given by

$$Z = \sum_{n_{12}} h(n; n_1, n_{12}) \exp(-n_{12}\epsilon_{12}/kT), \tag{10}$$

where $\epsilon_{12} = \epsilon'_{12} - \frac{1}{2}(\epsilon'_{11} + \epsilon'_{22})$, and $\epsilon'_{11}, \epsilon'_{12}, \epsilon'_{22}$ are the interaction energies of the various nearest neighbour types. We can replace x_{12} by z [$= \exp(-\epsilon_{12}/kT)$] in (5) and take account only of the larger root; but our partition function is now given as the coefficient of $x_1^{n_1}$ in $A_1\lambda_1^n$, where λ_1 is the largest eigenvalue of the matrix

$$\begin{vmatrix} x_1 & x_1 z \\ z & 1 \end{vmatrix}. \tag{11}$$

This can be evaluated by the method of steepest descents. The required function can be put in the form†

$$\frac{1}{2\pi i}\int_C \frac{A_1\lambda_1^n\, dx_1}{x_1^{n_1+1}}, \tag{12}$$

where C is a contour encircling the origin. The major contribution to this integral for large n comes from the neighbourhood of the saddle point, $x_1 = \nu$, and following the usual argument we deduce that

$$F = -nkT\log\lambda_1 + n_1 kT\log\nu, \tag{13}$$

where
$$\frac{\nu}{\lambda_1}\frac{\partial\lambda_1}{\partial\nu} = \frac{n_1}{n} = \alpha_1,$$

λ_1 being the largest eigenvalue of $\begin{vmatrix} \nu & \nu z \\ z & 1 \end{vmatrix}$. This last result could also have been deduced by using the grand partition function and its properties (Fowler & Guggen-

---

[*] Equations of this type were used by Lassettre & Howe (1941) in their discussion of the problem.

[†] It is believed that difficulties which may arise from branch points are spurious, and due to neglect of the other root. The method seems similar to that used by Chang & Ho (1942).

heim 1939, ch. VI; Lassettre & Howe 1941; Montroll 1942). In fact, the grand partition function is given by

$$\zeta = \sum_{n_1, n_2, n_{12}} h(n;\, n_1, n_{12})\, \xi_1^{n_1} \xi_2^{n_2} z^{n_{12}}, \tag{14}$$

and is determined (except for a simple factor) by the largest eigenvalue of the matrix (11) with $x_1 = \xi_1 \xi_2^{-1}$.

It will be seen that all the problems discussed in this section will be solved if we know the largest eigenvalue of the matrix

$$\begin{vmatrix} \mu & \mu z \\ z & 1 \end{vmatrix}, \tag{15}$$

which depends on two variables μ and z. Similar considerations will apply to problems discussed in the remainder of this paper, and our main purpose will be to formulate the matrix appropriate to each problem. Either system of difference equations (2) or (9) is sufficient to determine this matrix, and we shall use (2) for convenience.

It is of some interest to compare the solution given by (13) with the corresponding solution given by (8) with $\mu = 1$. When the proportions of each component are equal the solutions are identical. This can be deduced from the relation $\mu^{-\frac{1}{2}}\lambda_1(\mu) = \mu^{\frac{1}{2}}\lambda_1(1/\mu)$ obtained by reversing the direction of the magnetic field, and is valid throughout the remainder of the paper. Solutions giving the energy E for the case $\alpha_1 = \frac{1}{4}$, $\alpha_2 = \frac{3}{4}$ are compared in figure 1.

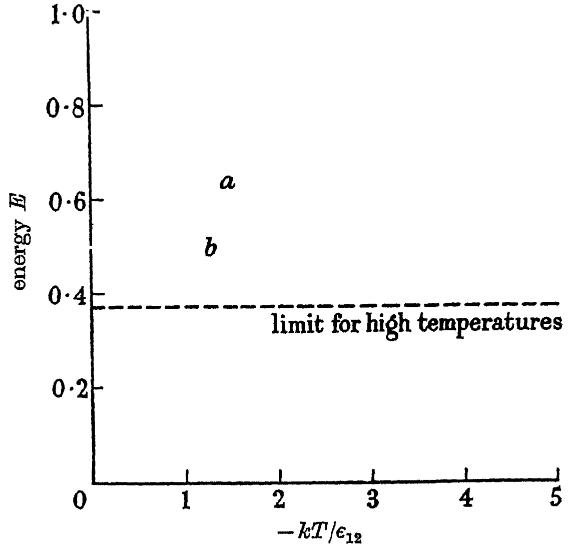


FIGURE 1. *a*. Solution (8). *b*. Solution (13).

3. TWO-DIMENSIONAL CASE; THE SCREW TRANSFORMATION

The method described in the previous section can be generalized to deal with a finite strip m units wide. A chain of n tiers contains mn units, and we must divide our configuration function $g(mn;\, n_1, n_{12})$ into mutually exclusive classes $g_{a_1,\ldots,a_m}(mn;\, n_1, n_{12})$ each corresponding to a given configuration $(a_1, \ldots, a_m)$ of the final tier. Each of $a_1, \ldots, a_m$ can be 1 or 2, so that there are 2^m such classes. We obtain a system of difference equations relating the $g_{a_1,\ldots,a_m}(mn+m;\, n_1, n_{12})$ with the $g_{a_1,\ldots,a_m}(mn;\, n_1, n_{12})$, and hence determine the partition function per tier as the largest eigenvalue λ_m of a matrix T_m of order 2^m. Some slight reduction in the order of T_m is possible because of symmetry considerations. For large m, if λ is the partition function per single unit, λ_m is effectively equal to λ^m. Results based on this approach have been deduced by other authors (Kramers & Wannier 1941; Montroll 1941, 1942; Lassettre & Howe 1941). The matrix T_m, however, is complicated to write down for all but the smallest m, and we show now that it can be put in the form $A_m^{m-1} B_m$, where A_m and B_m are matrices of simple characteristic structure. This reduction is very similar to the screw transformation introduced by Kramers & Wannier, but it shows clearly the relation between the matrices T_m and A_m.

Instead of adding a whole tier m units wide let us add one unit at a time; when we complete a tier we start from the bottom on the next tier. Our configuration function may now be written $g(p;\, n_1, n_{12})$, where p is the total number of units in the system. We divide this into 2^m mutually exclusive classes $g_{a_1,\ldots,a_m}(p;\, n_1, n_{12})$ each of which is characterized by the configuration $(a_1, \ldots, a_m)$ of the last m units added (figure 2).

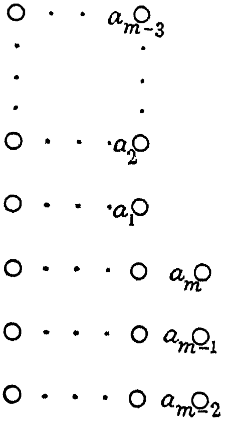


FIGURE 2

For convenience of writing we denote the function $g_{a_1,\ldots,a_m}(p;\, n_1, n_{12})$ by $\{a_1, \ldots, a_m\}(p;\, n_1, n_{12})$. It is easy to see that in changing the number n_{12} by the addition of one unit only a_1 and a_m play a part. Hence as long as p is not a multiple of m we deduce the following four equations:

$$\left.\begin{aligned}
&\{a_2, \ldots, a_{m-1}\,11\}(p+1;\, n_1, n_{12}) \\
&\quad = 2\alpha_1\{1\,a_2, \ldots, a_{m-1}\,1\}(p;\, n_1-1, n_{12}) + 2\alpha_1\{2\,a_2, \ldots, a_{m-1}\,1\}(p;\, n_1-1, n_{12}-1), \\
&\{a_2, \ldots, a_{m-1}\,12\}(p+1;\, n_1, n_{12}) \\
&\quad = 2\alpha_2\{1\,a_2, \ldots, a_{m-1}\,1\}(p;\, n_1, n_{12}-2) + 2\alpha_2\{2\,a_2, \ldots, a_{m-1}\,1\}(p;\, n_1, n_{12}-1), \\
&\{a_2, \ldots, a_{m-1}\,21\}(p+1;\, n_1, n_{12}) \\
&\quad = 2\alpha_1\{1\,a_2, \ldots, a_{m-1}\,2\}(p;\, n_1-1, n_{12}-1) + 2\alpha_1\{2\,a_2, \ldots, a_{m-1}\,2\}(p;\, n_1-1, n_{12}-2), \\
&\{a_2, \ldots, a_{m-1}\,22\}(p+1;\, n_1, n_{12}) \\
&\quad = 2\alpha_2\{1\,a_2, \ldots, a_{m-1}\,2\}(p;\, n_1, n_{12}-1) + 2\alpha_2\{2\,a_2, \ldots, a_{m-1}\,2\}(p;\, n_1, n_{12}).
\end{aligned}\right\} \tag{16}$$

We may order the configurations $(a_1, \ldots, a_m)$ by analogy with the scale of two; it is convenient for this purpose to label the first row and column as zero. We obtain a matrix of order 2^m in which the (ij) coefficient is zero except when $j = [\tfrac{1}{2}i]$ or $2^{m-1} + [\tfrac{1}{2}i]$, where $[x]$ denotes the integral part of x. This is a matrix of characteristic shape which can conveniently be referred to as a duo-diagonal matrix. As an example the matrix A_4 is

$$\left|\begin{array}{cccccccccccccccc}
\mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \mu z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \mu z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \mu z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right|, \tag{17}$$

using the notation of the previous section. We must now see what happens when p is a multiple of m. In this case only a_1 plays a part in determining the change of n_{12}. We obtain a system of equations analogous to (16) giving rise to a duo-diagonal matrix B_m. The non-zero elements in the first four rows which completely determine the matrix are $(\mu, \mu z)\,(z, 1)\,(\mu, \mu z)\,(z, 1)$ respectively instead of $(\mu, \mu z)\,(z^2, z)\,(\mu z, \mu z^2)$ $(z, 1)$. It is easy to see that the relation between the $\{a_1, \ldots, a_m\}\,(nm+m)$ and the $\{a_1, \ldots, a_m\}\,(nm)$, which gives rise to T_m, can be obtained by successive applications of systems such as (16), so that

$$T_m = A_m^{m-1} B_m. \tag{18}$$

We shall now assume that as $m \to \infty$, the mth root of the largest eigenvalue of $A_m^{m-1} B_m$ tends to the largest eigenvalue, λ, of A_m. The mathematical derivation of this result appears to be quite difficult. But on physical grounds it seems reasonable to suppose that B_m corresponds to an end effect which becomes negligible as $m \to \infty$. Hence we shall identify λ, which is the limit as $m \to \infty$ of the largest eigenvalue of A_m, with the partition function per single unit.

The system of equations (16) can easily be modified to deal with the case when there are different interactions in the different directions. We must now consider two different sets of nearest neighbour types in the different directions. The non-zero elements in the first four rows of the resulting matrix are $(\mu, \mu z)\,(zz', z')$, $(\mu z', \mu zz')$ and $(z, 1)$ respectively, where z refers to the horizontal and z' to the vertical interactions. This case is of some interest since the complete solution in the absence of a magnetic field given by Onsager (1944) applies to it, and it may help in establishing the properties of these duo-diagonal matrices. It is worth observing that the matrix B_m merely corresponds to no interaction in the vertical direction ($z' = 1$), and its largest eigenvalue should correspond to the one-dimensional partition function.

It is interesting to look at the case when each unit is capable of more than two configurations, say r different configurations. All that we have done so far can easily be generalized; the relation between the ferromagnetic and order-disorder problems is given by the generalization of equations such as (13), but the matrix corresponding to (15) which determines the complete solution of these problems, now depends on $(r-1)$ parameters of μ type, and on the temperature through $r(r-1)/2$ parameters of z type. The matrix A_m is of order r^m and is determined by a system of r^2 equations such as (16). It is an r-diagonal matrix, that is the (ij) coefficient is zero except when

$$j = [i/r],\ [i/r]+r^{m-1},\ [i/r]+2r^{m-1},\ \ldots,\ [i/r]+(r-1)\,r^{m-1}.$$

Where there is no magnetic field in the ferromagnetic problem the duo-diagonal matrix admits of groups of symmetric and antisymmetric eigenvalues on interchanging 1 and 2. Confining our attention only to the symmetric group we derive the V-matrix of Kramers & Wannier of order 2^{m-1}. Actually we first obtain a V-matrix with rearranged elements, but if we order our configurations according to $(a_m, \ldots, a_1)$ instead of $(a_1, \ldots, a_m)$ we derive a matrix which is effectively identical with theirs.

4. THE THREE-DIMENSIONAL TRANSFORMATION

If we consider n tiers, each tier consisting of an $l \times m$ rectangle, in a manner analogous to that described at the beginning of § 3, we obtain a matrix $T_{l,m}$ of order 2^{lm} whose largest eigenvalue, $\lambda_{l,m}$, is the partition function per tier for the block. We are interested in large l and m, and if λ is the partition function per single unit $\lambda_{l,m}$ is effectively equal to λ^{lm}. We now show that $T_{l,m}$ can be decomposed into a product of matrices, and it is reasonable to assume that λ is the limit as $l, m \to \infty$ of the largest eigenvalue of a duo-diagonal matrix. Let us again build up our system by adding one unit at a time, moving horizontally to complete each row, then

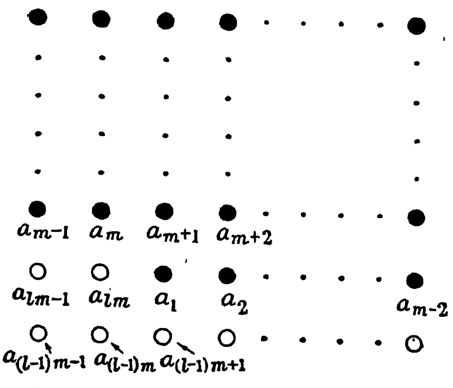


FIGURE 3. Shaded units refer to previous tier.

starting at the beginning of a new row until a tier is completed, then starting at the bottom of a new tier. We divide our configuration function into mutually exclusive classes as before, characterizing each class by the last lm units,

$$(a_1, \ldots, a_m, a_{m+1}, \ldots, a_{2m}, \ldots, a_{lm}) \quad \text{(figure 3)}.$$

It will be noticed that in adding a new unit only a_1, $a_{(l-1)m+1}$ and a_{lm} play a part. Hence we obtain a system of eight equations analogous to (16). We first deal with the case when the added unit is not a member of the first row or column. The other cases need special consideration although by assuming different interactions in the different directions, and then putting the appropriate interactions equal to zero, we can deduce them all from one set of equations. Let n_{12}, n'_{12}, n''_{12} refer to the three directions. For simplicity of writing only changes from n_1, n_{12}, n'_{12}, n''_{12} will be recorded; thus $(p; n'_{12}-1)$ is meant to indicate $(p; n_1, n_{12}, n'_{12}-1, n''_{12})$. Then

$$\left.\begin{aligned}
&\{a_2, \ldots, a_{(l-1)m}\, 1\, a_{(l-1)m+2}, \ldots, a_{lm-1}\, 11\}(p+1)\\
&\quad = 2\alpha_1\{1\, a_2, \ldots, a_{(l-1)m}\, 1, \ldots, a_{lm-1}\, 1\}(p; n_1-1)\\
&\qquad + 2\alpha_1\{2\, a_2, \ldots, a_{(l-1)m}\, 1, \ldots, a_{lm-1}\, 1\}(p; n_1-1, n''_{12}-1),\\
&\{a_2 \quad \cdots \quad 1 \quad \cdots \quad 12\}(p+1)\\
&\quad = 2\alpha_2\{1 \quad \cdots \quad 1 \quad \cdots \quad 1\}(p; n_{12}-1, n'_{12}-1, n''_{12}-1)\\
&\qquad + 2\alpha_2\{2 \quad \cdots \quad 1 \quad \cdots \quad 1\}(p; n_{12}-1, n'_{12}-1),\\
&\{a_2 \quad \cdots \quad 1 \quad \cdots \quad 21\}(p+1)\\
&\quad = 2\alpha_1\{1 \quad \cdots \quad 1 \quad \cdots \quad 2\}(p; n_1-1, n_{12}-1)\\
&\qquad + 2\alpha_1\{2 \quad \cdots \quad 1 \quad \cdots \quad 2\}(p; n_1-1, n_{12}-1, n''_{12}-1),\\
&\{a_2 \quad \cdots \quad 1 \quad \cdots \quad 22\}(p+1)\\
&\quad = 2\alpha_2\{1 \quad \cdots \quad 1 \quad \cdots \quad 2\}(p; n'_{12}-1, n''_{12}-1)\\
&\qquad + 2\alpha_2\{2 \quad \cdots \quad 1 \quad \cdots \quad 2\}(p; n'_{12}-1),\\
&\{a_2 \quad \cdots \quad 2 \quad \cdots \quad 11\}(p+1)\\
&\quad = 2\alpha_1\{1 \quad \cdots \quad 2 \quad \cdots \quad 1\}(p; n_1-1, n'_{12}-1)\\
&\qquad + 2\alpha_1\{2 \quad \cdots \quad 2 \quad \cdots \quad 1\}(p; n_1-1, n'_{12}-1, n''_{12}-1),\\
&\{a_2 \quad \cdots \quad 2 \quad \cdots \quad 12\}(p+1)\\
&\quad = 2\alpha_2\{1 \quad \cdots \quad 2 \quad \cdots \quad 1\}(p; n_{12}-1, n''_{12}-1)\\
&\qquad + 2\alpha_2\{2 \quad \cdots \quad 2 \quad \cdots \quad 1\}(p; n_{12}-1),\\
&\{a_2 \quad \cdots \quad 2 \quad \cdots \quad 21\}(p+1)\\
&\quad = 2\alpha_1\{1 \quad \cdots \quad 2 \quad \cdots \quad 2\}(p; n_1-1, n_{12}-1, n'_{12}-1)\\
&\qquad + 2\alpha_1\{2 \quad \cdots \quad 2 \quad \cdots \quad 2\}(p; n_1-1, n_{12}-1, n'_{12}-1, n''_{12}-1),\\
&\{a_2 \quad \cdots \quad 2 \quad \cdots \quad 22\}(p+1)\\
&\quad = 2\alpha_2\{1 \quad \cdots \quad 2 \quad \cdots \quad 2\}(p; n''_{12}-1)\\
&\qquad + 2\alpha_2\{2 \quad \cdots \quad 2 \quad \cdots \quad 2\}(p).
\end{aligned}\right\} \quad (19)$$

The matrix $A_{l,m}$ derived from this set of equations is duo-diagonal and of order 2^{lm}. The non-zero elements of any row $(a_1, a_2, \ldots, a_{lm})$ are completely determined by the values of $a_{(l-1)m}$, a_{lm-1}, a_{lm}. We thus have eight different cases corresponding to (111), (112), (121), (122), (211), (212), (221) and (222), and these are given respectively by $(\mu, \mu z'')$ $(zz'z'', zz')$ $(\mu z, \mu zz'')$ $(z'z'', z')$ $(\mu z', \mu z'z'')$ (zz'', z) $(\mu zz', \mu zz'z'')$ $(z'', 1)$. The matrix $B_{l,m}$ corresponding to the first unit of any row other than the first is obtained by putting $z = 1$; the matrix $C_{l,m}$ corresponding to any unit of the first row other than the first unit is obtained by putting $z' = 1$; and finally $D_{l,m}$ corresponding to the first unit of the first row is obtained by putting $z = z' = 1$. $B_{l,m}$ and $C_{l,m}$ correspond to two-dimensional problems, and $D_{l,m}$ to a one-dimensional problem. It will be seen that $T_{l,m}$ can be put in the form $(A_{l,m}^{m-1} B_{l,m})^{l-1} C_{l,m}^{m-1} D_{l,m}$. Letting l tend to infinity, an assumption analogous to that of § 3 indicates that the lth root of $\lambda_{l,m}$ tends to the largest eigenvalue of $A_{l,m}^{m-1} B_{l,m}$. Now letting $m \to \infty$ we should expect the lmth root of $\lambda_{l,m}$ to tend to the largest eigenvalue of $A_{l,m}$. Hence we shall identify λ, the limit as $l, m \to \infty$ of the largest eigenvalue of $A_{l,m}$, with the partition function per single unit.

As a typical example the matrix $A_{2,2}$ for the case in which all the interactions are equal, is

$$\left|\begin{array}{cccccccccccccccc}
\mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
z^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \mu z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \mu z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \mu z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^3 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & z^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z^2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \mu z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \mu z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & z & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mu z^3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right|. \tag{20}$$

When there is no magnetic field present the problem can, as in the two-dimensional case, be reduced to a V-matrix of degree 2^{lm-1}.

5. OTHER NEIGHBOUR INTERACTIONS; APPLICATION TO STATISTICAL MECHANICS OF A SYSTEM OF IDENTICAL PARTICLES

It is known that methods similar to that described in the first paragraph of § 3 can theoretically be generalized to take account of interactions other than nearest neighbour. But in practice the result is so complicated that it does not seem to have been attempted except in the simplest linear case (Montroll 1942). A transformation of the type used in the previous sections is applicable, and shows that this problem can also be reduced to the determination of the largest eigenvalue of a duo-diagonal matrix. The matter will not be discussed in great detail as the argument is closely analogous to that used previously. Let us take as an example the two-dimensional problem. Our system is built up by adding one unit at a time as before. We first take account of the next nearest neighbour which is along the diagonal. To do this we characterize each class of configurations by means of the last $(m+1)$ units added $(a_1, \ldots, a_{m+1})$ (figure 4). Let $n_{12}^{(1)}$ denote the number of nearest, and $n_{12}^{(2)}$ the number of next nearest neighbours of type 1–2. Then in the addition of a new unit $n_{12}^{(1)}$ and $n_{12}^{(2)}$ will only be affected by the values of a_2, a_{m+1} and a_1, a_3 respectively. Hence we can write down a system of sixteen equations similar to (19) which determine a duo-diagonal matrix of order 2^{m+1}. There are end effects, as before, but we assume that these can be neglected in the limiting case of large m. If we wanted to take account of third nearest neighbours we should require $(a_1, \ldots, a_{2m})$ (figure 5). The members which are effective in the addition of a new unit are a_{m+1}, a_{2m}; a_m, a_{m+2}; and a_1, a_{2m-1}. Hence the matrix appropriate to this case is of order 2^{2m} and there are sixty-four determining equations. Similarly for any other interactions. Although the number of determining equations rapidly becomes large, the actual process of finding the non-zero terms of any given row of the matrix is relatively simple.

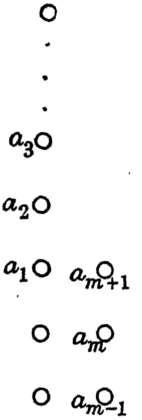


FIGURE 4

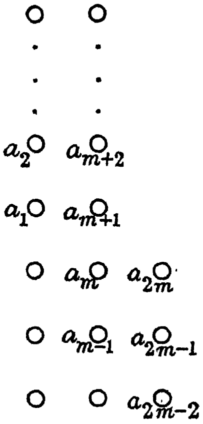


FIGURE 5

Since interactions other than nearest neighbour can be taken into account, one can formally apply the theory to the statistical mechanics of a system of identical particles. The exact problem in this case involves continuous variables instead of a discrete lattice structure, and any state of the system, solid, gaseous, etc., can be described by means of the distribution function $n(\rho)\,d\rho$ of the number of neighbours of a given particle within a distance ρ and $\rho+d\rho$. If this distribution function is given, the energy of the system can be written down immediately. The basic problem is the determination of the entropy of the system for any distribution function, or the number of configurations $H[n(\rho)\,d\rho]$ corresponding to the same distribution function $n(\rho)\,d\rho$. As an approximation to the exact problem we shall use a lattice structure as a co-ordinate mesh, the two configurations of each unit corresponding to a particle or no particle. This is equivalent to approximating to the continuous distribution $n(\rho)\,d\rho$ by a grouped series $[n_1(\rho_1), n_2(\rho_2), \ldots n_r(\rho_r)]$.

The use of a co-ordinate mesh is an approximation which is sufficiently near to the exact solution if the mesh is sufficiently fine. In order to obtain a reasonable physical approximation we must use enough neighbour interactions to give some idea of the shape of the inter-molecular potential-energy curve. Let us illustrate the method by considering the one-dimensional problem in which r neighbours are taken into account. Only 1–1 neighbour types give any interaction, and our problem is to determine the average number of configurations containing $n_{11}^{(1)}, n_{11}^{(2)}, \ldots, n_{11}^{(r)}$ neighbours of each kind respectively. Proceeding as before, using the same notation, and characterizing each class of configurations by means of the last r units added (figure 6) we obtain

$$\left.\begin{array}{l} \{a_2, \ldots, a_r\, 1\}(n+1) \\ \quad = 2\alpha_1\{1\, a_2, \ldots, a_r\}(n;\, n_{11}^{(1)}-1, n_{11}^{(k)}-1, \ldots, n_{11}^{(l)}-1) + 2\alpha_1\{2\, a_2, \ldots, a_r\}(n;\, n_{11}^{(k)}-1, \ldots, \\ n_{11}^{(l)}-1), \{a_2, \ldots, a_r\, 2\}(n+1) \\ \quad = 2\alpha_2\{1\, a_2, \ldots, a_r\}(n) \qquad\qquad\qquad\qquad + 2\alpha_2\{2\, a_2, \ldots, a_r\}(n), \end{array}\right\} \tag{21}$$

where $n_{11}^{(k)}, \ldots, n_{11}^{(l)}$ are easily determined from $a_2, \ldots, a_r$. The system (21) gives rise to a duo-diagonal matrix S_r of order 2^r. To take a particular case S_3 is

$$\left|\begin{array}{cccccccc} \mu z_{111} & 0 & 0 & 0 & \mu z_{211} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \mu z_{112} & 0 & 0 & 0 & \mu z_{212} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \mu z_{121} & 0 & 0 & 0 & \mu z_{221} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \mu z_{122} & 0 & 0 & 0 & \mu z_{222} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right|, \tag{22}$$

where $z_{111} = z^{(1)}z^{(2)}z^{(3)}$, $z_{211} = z^{(1)}z^{(2)}$, $z_{112} = z^{(2)}z^{(3)}$, $\ldots$; $z^{(r)} = \exp(-\epsilon^{(r)}/kT)$ and $\epsilon^{(r)}$ is the interaction energy of the rth neighbours. It is believed that quite a crude approximation ($r = 4$ or 5) will begin to give results of physical interest.

FIGURE 6

When we consider the problem in two or three dimensions an infinite duo-diagonal matrix is required, however many interactions are taken into account, and the number of interactions specifies the number of determining equations of this matrix, as in the first paragraph of this section. To obtain even a crude approximation to the shape of the intermolecular potential-energy curve the number of determining equations becomes large but in most practical cases the parameter μ in (22) may be regarded as small, and it may be possible to get an expansion in powers of μ. It is hoped to investigate this point further.

6. OTHER LATTICE STRUCTURES

So far we have considered only simple cubic lattices. But a transformation of the type used can be applied to any lattice, since this can always be constructed by the addition of single units. Hence any problem of this kind can be reduced to the determination of the largest eigenvalue of a product of duo-diagonal matrices. But the method is most powerful when with the exception of end effects all these constituent matrices are the same, so that the addition of a single unit is an operation of identical character for all units. Then the problem reduces to finding the largest eigenvalue of a single duo-diagonal matrix as in the simple cubic case.

As an example we take the face-centred cubic structure, since in this case one would expect the nearest neighbour approximation to be a good one. We must take the case in which the numbers of rows and columns are odd, because then the addition of a new unit is described by the same set of equations for all but the end units. Let the number of rows be $(2l+1)$ and the number of columns $(2m+1)$. Then to characterize a class of configurations we need $(a_1, a_2, \ldots, a_{(l+1)(2m+1)})$ (figure 7) so that our matrix will be of order $2^{(l+1)(2m+1)}$. But it is clear from figure 7 that only six units a_1, a_{m+1}, a_{m+2}, a_{2m+1}, $a_{2lm+lm+1}$, $a_{2lm+lm+2}$ affect n_{12} in the addition of a new unit. Hence there are sixty-four determining equations. It is quite an easy matter to write down the non-zero terms in any row of the matrix.

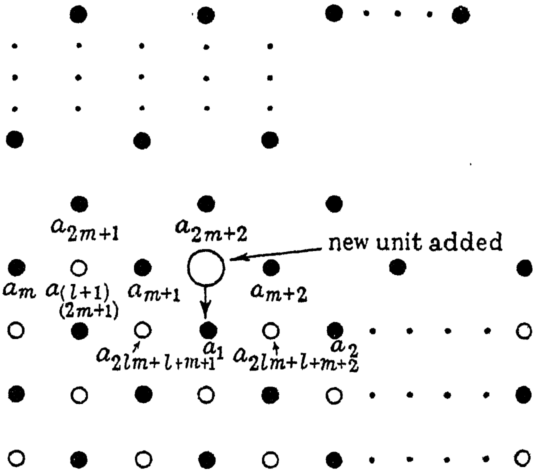


FIGURE 7. Shaded units denote previous tier.

The author is indebted to Dr A. R. Miller for drawing his attention to this problem, to Professor R. E. Peierls for commenting on the manuscript, and to the Department of Scientific and Industrial Research for a Senior Research Award.

REFERENCES

Bethe, H. A. 1935 *Proc. Roy. Soc.* A, **150**, 552.
Chang, T. S. & Ho, O. C. 1942 *Proc. Roy. Soc.* A, **180**, 345.
Fowler, R. H. & Guggenheim, E. A. 1939 *Statistical thermodynamics*. Cambridge University Press.
Kramers, H. A. & Wannier, G. H. 1941 *Phys. Rev.* **60**, 252, 263.
Lassettre, E. N. & Howe, J. P. 1941 *J. Chem. Phys.* **9**, 747.
Montroll, E. W. 1941 *J. Chem. Phys.* **9**, 706.
Montroll, E. W. 1942 *J. Chem. Phys.* **10**, 61.
Onsager, L. 1944 *Phys. Rev.* **65**, 117.
Wannier, G. H. 1945 *Rev. Mod. Phys.* **17**, 50.

Neutron diffraction

By G. E. Bacon and J. Thewlis

Atomic Energy Research Establishment, Harwell

(*Communicated by H. W. B. Skinner, F.R.S.—Received* 3 *June* 1948)

The problem of neutron diffraction by crystals is treated by analogy with X-ray diffraction, consideration being given to the perfect crystal, the mosaic crystal and the powdered block.

The first part of the paper deals with a comparison between X-ray and neutron diffraction and it is shown that quantitatively the two are similar, apart from the case of the thick mosaic crystal where the very low values which are usually found for the true absorption of neutrons result in the integrated reflexion being largely independent of structure factor under practical conditions.

The second part deals with the practical problem of diffracting the collimated beam of thermal neutrons from an atomic pile. Again, perfect and mosaic crystals are considered and it is also shown that the powder method, which it is desirable to use when structure factors are to be determined, will be feasible if the number of pile neutrons which hit the monochromatizing crystal is greater than 10^5 per sec.

INTRODUCTION

A number of papers have recently appeared on the subject of neutron diffraction. These have dealt with various aspects of the nature of the reflexion by matter of beams of thermal neutrons, produced in atomic piles. On the whole, the papers have been written from the standpoint of nuclear physics. It has, therefore, been considered worth while to write this appreciation of the problems of neutron diffraction from the point of view of the worker in the field of crystal analysis. The paper is divided into two parts, part I dealing largely with the similarities and differences between the diffraction of X-rays from a theoretical point of view, and part II with the actual problem of diffracting the collimated beam of thermal neutrons which can be obtained from an atomic pile.

PART I. COMPARISON OF X-RAY AND NEUTRON DIFFRACTION

1. *Wave-length of thermal neutrons*

As will be considered in more detail later, the collimated beam of neutrons which can be obtained from an atomic pile consists largely of a beam of 'slow' or 'thermal' neutrons. The distribution of velocities among the neutrons in the beam follows approximately the Maxwellian curve appropriate to neutrons in thermal equilibrium at a temperature of the order of 100° C.

The wave-length λ appropriate to neutrons having the root-mean square velocity v at temperature T° absolute is given by the equations

$$\lambda = \frac{h}{mv} \quad \text{and} \quad \tfrac{1}{2}mv^2 = \tfrac{3}{2}k_0 T,$$

whence

$$\lambda^2 = \frac{h^2}{3mk_0 T},$$

where m is the neutron mass, h is Planck's constant and k_0 is Boltzmann's constant. For a temperature of 100° C this gives a wave-length of 1·33 Å. At 0° C the wave-length would be 1·55 Å. Neutrons which have a velocity at the peak of the velocity distribution curve will have energy equal to $k_0 T$ and hence a wave-length which is $\sqrt{\tfrac{3}{2}}$ greater than that calculated above. Their wave-length will, therefore, be 1·63 and 1·90 Å respectively for the two temperatures considered. The wave-lengths associated with these beams of thermal neutrons are, therefore, close to the X-ray wave-lengths used in crystal analysis.

2. *Diffraction by a unit cell*

The square of the amplitude of the neutron wave will determine the number of neutrons per sec. at a point, just as the square of the amplitude of the electromagnetic wave determines the X-ray intensity. It is well known that a free electron in an electromagnetic field of amplitude A_0 gives at a distance r a scattered amplitude equal to

$$A_0 \frac{e^2}{mc^2} \frac{1}{r} \sin\theta,$$

where θ is the angle between the direction of measurement and the direction of vibration of the incident electromagnetic wave. In particular, for the polarized component whose electric intensity is at right angles to the plane of the incident and diffracted beams, the scattered amplitude from the electron is

$$A_0 \frac{e^2}{mc^2} \frac{1}{r}.$$

Proceeding further to the case of an atom, we have, with the usual definition of f_0 the atomic scattering factor, that the scattered amplitude is

$$A_0 f_0 \frac{e^2}{mc^2} \frac{1}{r}.$$

In the case of neutron scattering, the effective scatterer is the nucleus. The scattering is spherically symmetrical and since it can be shown from wave-mechanical considerations that the expression for the scattered amplitude is of the form $A_0 y(1/r)$ we can at once express y in terms of the measured nuclear scattering cross-sections. The total energy scattered by a nucleus will be equal to

$$4\pi r^2 A_0^2 y^2 \frac{1}{r^2},$$

which equals $4\pi A_0^2 y^2$ per nucleus. The fraction of energy scattered per nucleus is therefore $4\pi y^2$ and hence the fraction of energy scattered per unit length of path will be $4\pi y^2 n$, where n is the number of nuclei per cm.^{3}.

If σ is the effective cross-section for scattering then it follows that the fraction of neutrons scattered per cm. of path is also equal to σn.

Hence $y = \sqrt{\left(\frac{\sigma}{4\pi}\right)}$ and the scattered amplitude from a nucleus, for unit incident amplitude, will be $\frac{1}{r}\sqrt{\left(\frac{\sigma}{4\pi}\right)}$. Thus, in place of the expression $f_0 \frac{e^2}{mc^2}$ for X-rays, we have $\sqrt{\left(\frac{\sigma}{4\pi}\right)}$ for neutrons, giving the scattered amplitude per atom for unit incident amplitude.

Before considering the amplitude of the wave scattered by a unit cell it is desirable to stress that the scattering cross-section used above will not necessarily be the same as the total scattering cross-section measured by nuclear physical methods. This arises from the possible presence of various isotopes of an element in the same crystal, from the existence of non-zero spin for many nuclei, from thermal vibrations and from inelastic scattering.

We will consider first the case of a random distribution of isotopes, other disturbing factors being ignored. Then the cross-section for scattering into the diffraction peaks is given by

$$\sigma_{\text{Bragg}} = 4\pi(\Sigma py)^2,$$

where p is the isotopic abundance, whereas the total scattering cross-section, which includes background scattering, is given by

$$\sigma_{\text{total}} = 4\pi\Sigma(py^2).$$

In these expressions the actual values of the y's, and therefore of σ, are those appropriate to the state of binding of the nuclei in the crystal, the cross-section of a bound atom being greater than that of the same atom when free by a factor $\left(\frac{A+1}{A}\right)^2$, where A is the mass number. Thus

$$\sigma_{\text{bound}} = \sigma_{\text{free}}\left(\frac{A+1}{A}\right)^2.$$

Unless the scattering amplitude y is the same for all isotopes, or the element is monoisotopic or nearly so, σ_{Bragg} will be less than σ_{total}, particularly if the sign of the scattering amplitude varies. This difference, like similar differences discussed below, will manifest itself as an increase in the background scattering.

Similarly, in the case of finite spin, the scattering amplitude may be different according as the spin vector of the scattering nucleus is oriented parallel or anti-parallel to the spin vector of the neutron; and again σ_{Bragg} will in general be less than σ_{total}. For nuclei of zero spin (e.g. isotopes of even atomic mass above N^{14}) the spin effect will, of course, be absent.

Thermal vibrations will have the effect of reducing the diffracted intensity, as in the case of X-rays; and it has, in fact, been shown by Weinstock (1944) that the Debye-Waller formula holds equally well for neutrons.

The question of inelastic scattering is much more important in the case of neutron diffraction than in that of X-ray diffraction. A thermal neutron will lose an appreciable fraction of its energy in exciting a quantum of lattice energy, the frequency of vibration being of the same order (10^{13} per sec.) in each case; whereas the loss of energy for an X-ray quantum of the same wave-length is negligible, as the X-ray frequency is of the order of 10^{18} per sec. Weinstock (1944) calculated that the ratio of the inelastic to the 'free' elastic cross-section should increase with temperature and in the one numerical example worked out, that of polycrystalline iron for 300° K neutrons, he found that this ratio rose from a value of 0·006 at 0° K to 0·192 at 1000° K.

In the present paper, therefore, it should be understood that σ refers to σ_{Bragg} which will in general be less than σ_{total}. In some cases, e.g. hydrogen and lead, it is very much less, but in many others there is evidence (Fermi & Marshall 1947) that the dependence of scattering amplitude on spin orientation and isotopic constitution is not very marked. The question of binding, isotope and spin effects is treated more fully in a paper by Wollan & Shull (1948).

With this understanding of the interpretation of σ it follows that the amplitudes of the waves scattered by a unit cell in a direction specified by the Miller indices hkl are, for unit incident amplitude,

$$F_X = \frac{e^2}{mc^2}\Sigma f_0 \exp\left[2\pi i\left(\frac{hx}{a}+\frac{ky}{b}+\frac{lz}{c}\right)\right] \quad \text{for X-rays}$$

and

$$F_N = \Sigma \sqrt{\left(\frac{\sigma}{4\pi}\right)} \exp\left[2\pi i\left(\frac{hx}{a}+\frac{ky}{b}+\frac{lz}{c}\right)\right] \quad \text{for neutrons.}$$

In the above expression a, b, c are the sides of the unit cell and the summation is made over the various atomic positions x, y, z.

It should be noted that the above quantity F_X differs from the usually defined structure factor F for X-rays by including the factor e^2/mc^2. An examination of quantitative values shows that these two expressions are of the same order of magnitude for many atoms. For most atoms the value of σ lies between 2 and 8×10^{-24} sq.cm., and an average value is about 4×10^{-24}. Consequently $\sqrt{(\sigma/4\pi)}$ averages about 6×10^{-13}. It will be convenient to use MgO, for which $a = 4{\cdot}2$ Å, as a substance for which the X-ray and neutron reflexions may be compared. Goldberger & Seitz (1947) give $\sigma_{\text{Mg}} = \sigma_0 = 4 \times 10^{-24}$ cm.², and hence $\sqrt{(\sigma/4\pi)}$ equals 6×10^{-13} cm., whereas for X-rays $f_0(e^2/mc^2)$ is 14×10^{-13} cm. for magnesium and 25×10^{-13} cm. for oxygen for the 200 reflexion. Thus within a factor of three or four the scattering factors of magnesium oxide for neutrons and X-rays will be the same.

It should be noted here that it is the fact, referred to above, that the value of σ is of the same order for all nuclei which gives neutron diffraction particular value for certain structure determinations. This comparative independence of scattering power on atomic number is of course in sharp contrast to the continual increase with atomic number of the scattering power of atoms for X-rays.

3. *Amplitude of reflexion by a crystal plane*

We can now proceed to the calculation of the amplitude of radiation reflected by a single net plane of the crystal. Following the treatment given on p. 559 of the *International Tables for the Determination of Crystal Structures* (1935), the ratio of the amplitude of the reflected and incident beams, neglecting polarization, is

$$Nd\frac{\lambda}{\sin\theta}F_X$$

for X-rays, where N is the number of unit cells per unit volume, d is the spacing and θ the Bragg angle. For the 200 reflexion of MgO and using Cu $K\alpha$ radiation, this ratio is equal to $1\cdot7\times10^{-4}$. For neutrons, the ratio is $Nd\frac{\lambda}{\sin\theta}F_N$, which for the same conditions and equal wave-length has a value of $0\cdot6\times10^{-4}$.

Since these ratios are of the same order it follows that the classification of crystals into perfect and 'mosaic' will be on much the same lines in the case of neutrons as in the case of X-rays.

4. *Calculation of Q*

We proceed to the calculation of Q, the expression which appears in most formulae for diffraction intensities. In particular the integrated reflexion for a small element of crystal of volume δV bathed in radiation is $Q\,\delta V$, assuming that the element is so small that absorption and extinction in it are negligible.

For X-rays

$$Q=\frac{N^2}{\sin 2\theta}F_X^2\lambda^3,$$

which for MgO 200 has the value $0\cdot2$ cm.$^{-1}$ for Cu $K\alpha$ radiation.

For neutrons

$$Q=\frac{N^2}{\sin 2\theta}F_N^2\lambda^3$$

which for MgO 200 and the same wave-length has a value of $0\cdot02$ cm.$^{-1}$.

5. *Reflecting power of perfect crystal*

For the perfect crystal the integrated reflexion is equal to $\frac{8}{3}s$, the area under the Darwin curve, where s is equal to

$$NF_X\frac{\lambda^2}{\pi\sin 2\theta}\quad\text{for X-rays,}$$

and

$$NF_N\frac{\lambda^2}{\pi\sin 2\theta}\quad\text{for neutrons.}$$

Again, taking the case of the 200 reflexion of MgO for $\lambda = 1\cdot54$ Å, the values of $\frac{8}{3}s$, are $6\cdot4\times10^{-5}$ (12 sec.) and 2×10^{-5} (4 sec.) for X-rays and neutrons respectively.

6. *Reflecting power of large mosaic crystal*

An important distinction between the reflexion of X-rays and neutrons appears when the case of the large mosaic crystal is considered. For a large slab of mosaic crystal the reflectivity for X-rays for the crystal planes which are parallel to the surface of the specimen is taken as $Q/2\mu$ if secondary extinction is neglected entirely or $Q/2(\mu+gQ)$ if secondary extinction is assumed small but a correction is made for it. In these expressions, g is a constant depending on the extent of crystal perfection and μ is the linear absorption coefficient for true absorption. Secondary extinction cannot be ignored or allowed for in this manner unless the power loss due to diffraction is negligible or small respectively compared with the loss due to true absorption (i.e. absorption not due to scattering). This is, in fact, the case for X-rays, where the mass scattering coefficient is of the order of 0·2 for most elements whereas the mass absorption coefficient is much larger, e.g. about 12 and 40 for oxygen and magnesium respectively for Cu $K\alpha$ radiation.

In the case of neutrons, however, the loss of energy due to scattering is very much greater than that due to true absorption for most elements. For example, the linear coefficient of true absorption for MgO for neutrons has a value of about 0·02 in comparison with a value for X-rays of about 100. It is only when elements such as cadmium, which is a relatively intense absorber of neutrons, are present that the linear absorption coefficient has values of the same order as those found for most substances with X-rays.

The above expression for the integrated reflexion of a mosaic crystal will not in general hold for neutrons owing to the small value of the true absorption coefficient relative to the scattering coefficient. In most cases it is necessary to use a general expression such as that given as equation (4·24) in § IV. 3 of Zachariasen (1944). In this treatment it is assumed that the mosaic blocks are oriented according to an error function $W(\Delta)$, where $W(\Delta)\,d\Delta$ is the number of blocks whose normals make angles between Δ and $\Delta+d\Delta$ with the mean normal. The expression for $W(\Delta)$ is

$$W(\Delta) = \frac{1}{\eta\sqrt{2\pi}} \exp\left[-\Delta^2/2\eta^2\right],$$

where η is the standard deviation. The actual form which the expression for integrated reflexion takes in practical cases will be the subject of a further paper. However, the following general conclusions are stated here. If μ is large, as for the X-ray case, then the integrated reflexion can be taken as $Q/2\mu$ for all practical thicknesses of crystal. When μ is very small, as for neutrons, there is a region of thickness t (of the order of a fraction of a millimetre) for which the integrated reflexion is given by $Qt/\sin\theta$. With increase of t we approach a range of thicknesses likely to be met with in practice where secondary extinction is of predominant importance. For these thicknesses we reach what is, at first sight, the startling conclusion that the integrated reflexion is equal to about 4η and is little influenced by Q. It can be seen from the shape of the Gaussian curve that this quantity 4η is what might be called the base width of the mosaic spread. Clearly every significant part of the mosaic spread is contributing totally to the integrated reflexion.

Finally, when t is made very large the integrated reflexion slowly increases to a limit greater than this for very small values of μ, as appreciable reflexion can then be obtained from crystal elements at the tail of the mosaic distribution curve. With continual reduction of μ this limit as t approaches infinity increases, unit reflectivity being obtained over an ever increasing angular range. If the absorption were zero and the crystal thick enough there would always be sufficient mosaic blocks, on the distribution assumed, to reflect all the beam at all inclinations.

As an example in the case of practical values of μ and t let the value of 4η, the base width of the mosaic distribution curve, be 4 min. and the width of the Darwin curve 4 sec. The integrated reflexion of the mosaic crystal would then be 60 times greater than that of the perfect crystal. In general, for moderately imperfect crystals, enhancement of 10 to 100 times may be expected for neutrons. With X-rays, where the relatively high absorption is all important, the enhancement is less. For example, with a very imperfect rock-salt crystal for which η was about 450×10^{-5} (15 min.) the measured value of the integrated reflexion was (Zachariasen 1944) 27×10^{-5} compared with a $Q/2\mu$ value of 33×10^{-5} and a Darwin curve width of 4×10^{-5}. Thus, in this case, the enhancement of the mosaic crystal relative to the perfect crystal is about seven times for X-rays, whereas an enhancement of about 400 would be expected for neutrons.

Summarizing, it should be noted that the integrated reflexion of a 'thick' mosaic crystal for neutrons is, to a first order at least, independent of Q and hence the structure factor. For very thin crystals the integrated reflexion will be proportional to Q and hence to the square of the structure factor. With increase of thickness there is a transition from this proportionality to the region where the integrated reflexion is largely independent of Q and determined mainly by the width of the mosaic curve. In general, measurement of integrated reflexions of mosaic crystals for neutrons is a very difficult and unsatisfactory way of determining structure factors.

7. *Reflecting power of powdered block*

It will be shown that in contrast to the use of single crystals the use of a block of powdered crystal will enable us to determine values of structure factors.

The powdered crystal can be regarded as the limiting case of a mosaic crystal in which the orientation of the reflecting blocks has become completely random, instead of being limited by the width of a relatively narrow distribution curve. For such a random distribution it can be shown that the probability of finding a mosaic block oriented within a small range of $d\theta$ of the Bragg angle is equal to $\frac{1}{2}\cos\theta\, d\theta$ (see for example Compton & Allison 1935, p. 416). It then follows that for symmetrical reflexion by a block of thickness t the total power in the whole of the diffraction cone is equal to

$$\frac{Qp\cos\theta}{4\mu}(1-\exp[-2\mu t\,\mathrm{cosec}\,\theta])\,,$$

just as for X-rays, p being the number of co-operating planes. The fact that the orientation of the crystal blocks is purely random implies that secondary extinction need not be considered as such and we can regard the reduction of beam intensity in passing into the crystal as due entirely to a linear absorption coefficient which, for the majority of elements, will be equal to $\Sigma\sigma_s n$ summed over the various nuclei which comprise the substance. Here σ_s is the total cross-section for scattering which we have previously considered. For elements where true absorption is appreciable we replace σ_s by σ_t which allows for both scattering and absorption. In the case of magnesium oxide $\sigma_s = \sigma_t = 4 \times 10^{-24}$ sq.cm. for both magnesium and oxygen and n is equal to $4/(4{\cdot}2 \times 10^{-8})^3$ nuclei per cm.^{3} for each element. From these values it follows that μ is equal to 0·44. This value is so small that the term $e^{-2\mu t}\operatorname{cosec}\theta$ in the above expression is not negligible even when t is quite large. For example, with a slab of thickness $\frac{1}{2}$ cm. the complete expression in the bracket has a value of about $\frac{2}{3}$. In the X-ray case where μ has a value of about 100 the expression in the bracket is effectively equal to unity even for much smaller thicknesses of crystal.

If the counter collects a length l of the circumference of the diffraction cone, which has a total circumference of $2\pi R\sin 2\theta$, where R is the distance of the counter slit from the crystal, then the reflecting power will be approximately

$$\frac{2}{3}\frac{Qpl}{16\mu R\sin\theta}.$$

This will be the ratio of the number of reflected neutrons per sec. entering the counter to the number of neutrons per sec. hitting the powdered crystal. In particular the value is proportional to Q and hence to the square of the structure factor, thus enabling the latter to be determined.

The above expression for the intensity of diffracted neutrons entering the counter was obtained for the case where the incident beam is reflected symmetrically from the surface of the powdered block. In a similar manner, an expression can be obtained for the case in which the incident beam is transmitted through the block. For this latter case, the expression can be shown to be

$$\frac{Qpl}{8\pi R\sin\theta}\exp\left[-\mu t\sec\theta\right] t\sec\theta.$$

This has a maximum value when the thickness t is equal to

$$\frac{1}{\mu\sec\theta}$$

and this maximum value is equal to

$$\frac{Qpl}{8\pi R\mu\sin\theta}\frac{1}{e}.$$

This value will be much the same as that for the reflexion method but the transmission method has the advantage that for small values of Bragg angle θ a smaller width of block is necessary for intercepting a given width of incident beam.

8. *Phase of scattered neutron wave*

In the case of X-rays there is always a phase difference of 180° between the incident wave and the resultant scattered wave from an atom. For neutrons this is again true for scattering by most nuclei but not for all. The only exceptions suggested at present are hydrogen, lithium, titanium and manganese which scatter with zero phase difference. The result of this is that for certain compounds of these elements the relative intensities of the orders of various spectra are quite different for neutrons from what they are for X-rays. In the case of LiF, for example, the odd orders of neutron spectra from the (111) planes are strong and the even orders weak. Fermi & Marshall (1947) report that the 222 reflexion is almost zero, from which it would follow that σ_{Li} and σ_{F} are numerically almost equal. If this were so it would be expected that for such a face-centred lattice structure all orders of reflexion from any plane parallel to one of the axes would be zero. However, Sturm (1947) reports much experimental work using the (200) planes of lithium fluoride, though it has been reported more recently by Wollan & Shull (1948) that this reflexion is very weak.

PART II. DIFFRACTION OF NEUTRON BEAM FROM ATOMIC PILE

9. *Collimated thermal neutron beam*

We now proceed to a determination of the effective reflectivity of a large crystal face for the collimated beam of thermal neutrons which can be obtained from an atomic pile. To a first approximation the neutrons can be considered to be in temperature equilibrium at a temperature somewhat higher than room temperature, probably 100° C, and their velocities will be distributed (approximately) according to a Maxwellian distribution. The collimator, which is some feet long, effectively withdraws from the atomic pile a representative fraction of these neutrons with their directions of motion limited within suitably small angular limits for the diffraction measurements. In order to reflect a sufficiently large number of neutrons for detection it is necessary to use a collimated beam with a cross-section of a few square inches.

This is achieved, while still maintaining a small angular divergence, by the use of a bundle of narrow collimators instead of a single collimator of large cross-section.

For neutrons in thermal equilibrium at temperature $T°$ absolute the number having velocities between v and $v+dv$ will, according to kinetic theory, be

$$k\,\mathrm{e}^{-\alpha m v^2} v^2\, dv,$$

where m is the neutron mass and k and α are constants for a given temperature. It can be shown that

$$\alpha = \frac{1}{2k_0 T},$$

where k_0 is Boltzmann's constant.

The number of neutrons with velocities between v and $v+dv$ hitting the crystal per sec., after passing down the collimator, we write as $\nu_v dv$ and it will be proportional to the product of v and the above expression, i.e.

$$\nu_v dv = k'\,\mathrm{e}^{-\alpha m v^2}\, v^3\, dv,$$

where k' is another constant.

The total number of neutrons, of all velocities, hitting the crystal will be

$$N_1 \equiv \int \nu_v \, dv$$

$$= k' \int e^{-\alpha m v^2} v^3 \, dv$$

integrated over all velocities, which can be shown to equal

$$\frac{k'}{2\alpha^2 m^2}.$$

Hence, eliminating k', we have

$$\nu_v \, dv = 2N_1 \alpha^2 m^2 \, e^{-\alpha m v^2} v^3 \, dv.$$

Expressing α as $1/2k_0 T$ as stated above, we have

$$\nu_v \, dv = 2N_1 \left(\frac{m}{2k_0 T}\right)^2 \exp\left[-\frac{\frac{1}{2}mv^2}{k_0 T}\right] v^3 \, dv.$$

Considering the distribution in terms of wave-length λ instead of velocity v and using the relation

$$\lambda = \frac{h}{mv}$$

the above reduces to $\nu_\lambda \, d\lambda = \dfrac{2N_1}{\lambda}\left(\dfrac{\epsilon}{k_0 T}\right)^2 \exp\left[-\dfrac{\epsilon}{k_0 T}\right] d\lambda,$

where $\epsilon = \frac{1}{2}mv^2$ is the energy of a neutron of velocity v or wave-length λ and $\nu_\lambda \, d\lambda$ is the number of neutrons with wave-length lying between λ and $\lambda + d\lambda$.

10. *Reflexion of collimated beam by single crystal*

We see from the above that the beam of neutrons emerging from the collimator to strike the crystal will be a beam of 'white' radiation having a spectrum defined by ν_λ as calculated. It is assumed that the crystal, when oriented with its surface making a particular angle with the beam, is sufficiently large to intercept the whole beam. Considering first of all a perfect crystal and assuming, for simplicity of argument, that the incident beam has no divergence, we can transform the normal Darwin R, θ curve into an R, λ curve for the particular Bragg angle θ for the incident beam using the fact that $\lambda = 2d \sin\theta$. The number of neutrons reflected per sec. for the wave-length range λ to $\lambda + d\lambda$ will be $R_\lambda \nu_\lambda \, d\lambda$ and the total number reflected will be $\int R_\lambda \nu_\lambda \, d\lambda$. Since $\lambda = 2d \sin\theta$ it follows that neutrons having a wave-length which differs by a small quantity $\Delta\lambda$ from the wave-length which undergoes Bragg reflexion will be incident at an angle which is $\Delta\lambda/2d \cos\theta$ different from their own Bragg angle. Consequently they will have a reflecting power appropriate to this angular displacement from the centre of the Darwin R, θ curve. Hence the area under the $R\lambda$ curve will be equal to $2d \cos\theta$ times the area under the $R\theta$ curve. Accordingly, the value of $\int R\nu_\lambda \, d\lambda$, which is the total number of reflected neutrons, will be equal to $\nu_\lambda \times 2d \cos\theta \times$ (area under Darwin curve), since ν_λ is effectively constant over the small range of λ for which appreciable reflexion takes place.

This expression will apply to any particular angular component of the divergent beam, and hence will apply to the divergent beam as a whole, taking ν_λ, as calculated above, to include neutrons in the whole of the beam. This assumes, of course, that the slit of the counter is sufficiently wide to catch the whole of the reflected beam and that the divergence of the beam is so small that ν_λ still remains effectively constant over the range of λ which undergoes reflexion.

As already stated, the area under the Darwin curve is equal to

$$\tfrac{8}{3}NF_N\frac{\lambda^2}{\pi\sin 2\theta},$$

so that the above expression for the total number of neutrons reflected per sec. by the perfect crystal becomes

$$\frac{2N_1}{\lambda}\left(\frac{\epsilon}{k_0T}\right)^2 e^{-\epsilon/k_0T}\,2d\cos\theta\,\frac{8}{3V}\,\frac{\lambda^2}{\pi\sin 2\theta}F_N,$$

where $V = 1/N$ is the volume of the unit cell. This expression reduces to

$$\frac{8}{3\pi}\frac{N_1}{V}\left(\frac{\epsilon}{k_0T}\right)^2\exp\left[-\frac{\epsilon}{k_0T}\right]4d^2F_N.$$

Although the energy at the peak of the thermal curve is k_0T, that at the peak of the distribution in the beam hitting the crystal can be shown to be $\tfrac{3}{2}k_0T$. Consequently, if the crystal is so oriented that the beam is incident at the Bragg angle for neutrons with velocities corresponding to the peak of the latter distribution then the total number of neutrons reflected per sec. N_2 will be

$$N_1\frac{8}{3\pi}\frac{1}{V}\left(\frac{3}{2}\right)^2 e^{-\frac{3}{2}}\,4d^2F_N,$$

which equals

$$\frac{1{\cdot}7d^2F_N}{V}N_1,$$

where N_1 is again the total number of neutrons of all velocities hitting the crystal per sec.

As an example we again take the 200 reflexion of MgO for which $a = 4{\cdot}2\times 10^{-8}$, $F_N = 48\times 10^{-13}$ and $d = 2{\cdot}1\times 10^{-8}$, from which the fraction of incident neutrons reflected is 5×10^{-5}. As discussed previously, the reflectivity for a mosaic crystal will be some 10 to 100 times greater than this. Hence the reflectivity of a mosaic crystal for a beam of thermal neutrons having a Maxwellian velocity distribution is of the order of 10^{-3} when the crystal is oriented at the Bragg angle for the peak of the beam distribution.

In addition to the reflected neutrons calculated as above there will be higher order reflexions, at the same Bragg angle, of wave-lengths which are sub-multiples of that considered. These wave-lengths will be considerably removed from the peak of the beam distribution curve and the corresponding intensities will consequently be low.

11. *Reflexion of monochromatic beam*

The beam of reflected neutrons coming off the fixed crystal considered above will be approximately monochromatic, with a wave-length corresponding to the peak of the beam distribution curve. The wave-length spread will depend on the mosaic spread of the crystal and the divergence of the incident beam. If this reflected beam is now allowed to fall on a second crystal we can then make measurements, for the particular wave-length concerned, of integrated intensity for the various orders of reflexion from the crystal planes parallel to the surface of the second crystal, by measuring the total number of neutrons E collected when the crystal is rotated through the Bragg angle with angular velocity ω radians per sec., the counter following at twice the angular rate to ensure that all the reflected neutrons are collected.

For a perfect crystal the integrated intensity will, as for the X-ray case, be equal to the area under the Darwin curve, i.e.

$$\frac{E\omega}{N_2} = \frac{8s}{3} = \frac{8}{3\pi}\frac{\lambda^2}{V\sin 2\theta}F_N.$$

In the case of the 200 reflexion for MgO

$$\frac{E\omega}{N_2} = 2\times 10^{-5}.$$

Allowing a factor of 25, say, for enhancement by a mosaic crystal we should have $Ew/N_2 = 5\times 10^{-4}$.

In this expression E is equal to the total number of neutrons which enter the counter during rotation. Experimentally the important quantity is the rate of entry of neutrons into the counter during the period of count, since this determines how easily the particular reflexion can be distinguished from the background count which is usually of the order of 15 per min. The count per sec. will depend on what total angle is swept through in the rotation. If this angle is ϕ radians then the time of count will be ϕ/ω sec. and the mean number of counts per sec. $E\omega/\phi$, which equals $5N_2/\phi\times 10^{-4}$ counts per sec. for the mosaic crystal. For $\phi = 0{\cdot}1$ radians, i.e. rotation through about 5 degrees, this would give $5\times 10^{-3}N_2$ counts per sec.

From the earlier discussion we saw that a fraction 10^{-3} of the neutrons N_1 incident on the first crystal were reflected by it. It follows therefore that if the whole of these $N_1/1000$ reflected neutrons hit the second crystal and if all the twice reflected neutrons are collected by the counter, then the mean number of counts per sec. when the second crystal is rotated through an angle ϕ radians about the Bragg position is of the order of $5N_1/\phi\times 10^{-7}$. Assuming a rotation through 5 degrees in order to allow amply for mosaic and divergent spread we should get a mean number of counts per sec. of the order of $5\times 10^{-6}N_1$ for the two MgO crystals using the 200 reflexion.

The longer the time of count the smaller will be the neutron intensity which can be distinguished from the background. For example, if we counted over a sufficiently long period to get 10,000 background counts then an increase of 2 % in count (due to crystal reflected neutrons) would be significant. If, however, we can only count for 1 min., for which there will be 15 background counts, then only an increase of

50 % or more is significant, which means eight 'crystal counts' per min. For ease and rapidity of observation it is desirable to have a crystal count of the same order as the background count. This will be the case if $5\times 10^{-6}N_1$, is not less than $\frac{1}{4}$, i.e. if N_1 is not less than 5×10^4. This, therefore, is the order of the total number of thermal neutrons which must be allowed to hit the first crystal if the particular integrated reflexion under consideration is to be easily observed with a two crystal spectrometer.

The above method of directly measuring the integrated reflexion by rocking through the reflecting position is not the only possible method of measurement and will not be the most sensitive way of detecting a reflexion. It is clear that if the beam hitting the second crystal was perfectly parallel and monochromatic then with the crystal set at the Bragg angle, the number of reflected neutrons would be equal to N_2, the number incident. Because of the divergence of the initial collimated beam this will not be the case in practice and the number of neutrons entering the counter in the optimum position will be less than N_2. It may, however, be desirable to obtain a measure of the integrated reflexion by taking a count at several stationary positions about the Bragg angle. Since, as we have already seen in part I, we are unable to relate this value of integrated reflexion at all easily to the value of the structure factor we shall not analyze this second method in any detail.

For determination of structure factors we resort to the use of a block of powdered crystal. Referring back to § 7 we have the value

$$\frac{2}{3}\frac{Qpl}{16\mu R\sin\theta}$$

as the ratio of the number of reflected neutrons per sec. entering the counter to the number of neutrons per sec. hitting the powdered crystal. Taking $p = 6$, $l = 5$ cm., $R = 50$ cm., $\mu = 0{\cdot}44$, $\sin\theta = 0{\cdot}4$ and $Q = 0{\cdot}02$ cm.$^{-1}$ for the 200 reflexion of MgO we have as the number of neutrons entering the counter per sec. $3\times 10^{-3}N_2$, where N_2 is the number of neutrons hitting the powdered crystal block. If we assume that the reflexion will be detectable if this number is as large as a background count of 15 per min., then we find that N_2 must be greater than about 100. From our previous assumptions this implies that N_1, the number of neutrons per sec. hitting the monochromatizing crystal, must be greater than 10^5. In practice the American workers report a value for N_2 of about 400 per sec. This should be sufficient to determine structure factors by the powder method for many substances.

12. *Applications of neutron diffraction*

The ability to provide a mono-kinetic beam of neutrons has obvious applications in nuclear physics, for example, in the determination of absorption cross-sections; but from the point of view of crystal structure analysis there are two main possible applications of neutron diffraction which suggest themselves. Both arise from the fact mentioned above that the scattering powers of all atoms for neutrons are roughly of the same order.

(a) *The determination of the positions of light atoms*

It is well known that it is very difficult or impossible to determine by X-ray analysis the positions of light atoms in a structure in which heavy atoms are also present, since the intensity of reflexion is approximately proportional to the square of the atomic number. In the case of neutron diffraction, however, since the scattering powers of all atoms are roughly of the same order, it should clearly be possible to determine the positions of the light atoms in the circumstances considered. In particular, it should be possible to determine the positions of hydrogen atoms, which would have an obvious importance in organic chemistry, although the relatively high background scattering from hydrogen might make it necessary to substitute deuterium in certain cases.

An example of this type of application of neutron diffraction has already been reported in an investigation of the structure of sodium hydride by Shull, Wollan, Morton & Davidson (1948).

(*b*) *The determination of complicated structures*

In complicated structures the use of neutron diffraction in combination with X-ray diffraction results gives us in effect a method of altering the relative scattering powers of the atoms in a structure, and sometimes the scattering phases, although the atoms themselves remain in the same configuration. The importance of this in structure determination is obvious.

In order to carry out the above work completely it will be necessary to learn more about the scattering powers of atoms for neutrons, and it will be of interest to study pure isotopes in certain instances. The problem of background scattering will arise, and the isotope, spin, thermal and inelastic effects will need to be disentangled. The question of the temperature dependence of inelastic scattering, and perhaps even that of the existence of isotopic order-disorder transformations, will merit attention, and it is to be expected that the results obtained will have significance not only for crystal structure but for nuclear physics. At the moment, the variation of scattering power from nucleus to nucleus appears to be a matter of chance, but with the possibility that is opened up of making accurate determinations of this scattering power for individual isotopes it is perhaps not too much to hope that neutron diffraction may help indirectly to throw light on the structure of the nucleus itself.

The discussion of the technique of neutron diffraction has been restricted to the use of collimated beams of large cross-section and detection by boron trifluoride filled counters, since these are the most satisfactory methods available at present. With improvements in photographic methods of neutron detection and possible increases in neutron beam intensity it may be possible to employ the types of beam and specimen familiar in X-ray analysis, including cylindrical powder specimens and, perhaps, even very small single crystals.

This paper is published by permission of the Director of the Atomic Energy Research Establishment. The authors are indebted to Professor Sir Lawrence Bragg, F.R.S., Dr Kathleen Lonsdale, F.R.S. and Dr H. W. B. Skinner, F.R.S. for valuable discussion and comment.

REFERENCES

Compton, A. H. & Allison S. K. 1935 *X-rays in theory and experiment.* New York: Van Nostrand.

Davidson, W. L., Morton, G. A., Shull, C. G. & Wollan, E. O. 1947 *Phys. Rev.* **72**, 168.

Fermi, E. & Marshall, L. 1947 *Phys. Rev.* **71**, 666.

Fermi, E., Sturm, W. J. & Sachs, R. G. 1947 *Phys. Rev.* **71**, 587.

Goldberger, M. L.. & Seitz, F. 1947 *Phys. Rev.* **71**, 294.

International Tables for the Determination of Crystal Structures 1935 Ann Arbor: Edwards Bros.

Shull, C. G. & Wollan, E. O. 1947 *Phys. Rev.* **72**, 168.

Shull, C. G., Wollan, E. O., Morton, G. A. & Davidson, W. L. 1948 *Phys. Rev.* **73**, 842.

Sturm, W. J. 1947 *Phys. Rev.* **71**, 757.

Weinstock, R. 1944 *Phys. Rev.* **65**, 1.

Wollan, E. O. & Shull, C. G. 1948 *Phys. Rev.* **73**, 830.

Zachariasen, W. H. 1944 *Theory of X-ray diffraction in crystals.* New York: Wiley.

Zinn, W. H. 1947 *Phys. Rev.* **71**, 752.

Some crystal-boundary phenomena in metals

By B. CHALMERS, *Atomic Energy Research Establishment, Harwell*

(*Communicated by E. N. da C. Andrade, F.R.S.—Received* 26 *June* 1948)

[Plate 1]

Techniques are described for preparing specimens of tin and of lead consisting of two or three crystals with controlled orientations. The direction of formation of the boundary between two crystals is shown to be dependent on the relative orientations of the crystal axes, and to be a result of a variation of the solid-liquid equilibrium temperature with the crystallographic characteristics of the solid surface in contact with the liquid. A qualitative explanation is advanced in terms of a theory of melting and freezing which envisages the simultaneous operation of a 'melting' process and a 'freezing' process which have equal rates at the equilibrium temperature. New observations of the 'macro-mosaic' effect are recorded, and it is shown that crystal boundaries can move at temperatures near the melting-point in the absence of plastic strain, if reduction of the area of the boundary is produced. It is also demonstrated that the specific free energies of all boundaries, except twin boundaries, in tin are equal, and do not depend on the relative orientations of the crystals.

1. INTRODUCTION

In an earlier paper (Chalmers 1940) a technique was described by means of which a crystal boundary could be formed in tin between two crystals with predetermined orientations. Attention was drawn to two phenomena which had not at that time been fully investigated. It was pointed out that if the orientations of the two crystals were not symmetrical with respect to the direction of growth of the crystals, the boundary tended to deviate from this direction. It was also pointed out that many of the crystals showed a secondary structure referred to as 'macro-mosaic'. In the present paper an analysis has been made of the conditions governing the direction of growth of the crystal boundary in tin and in lead, and some further

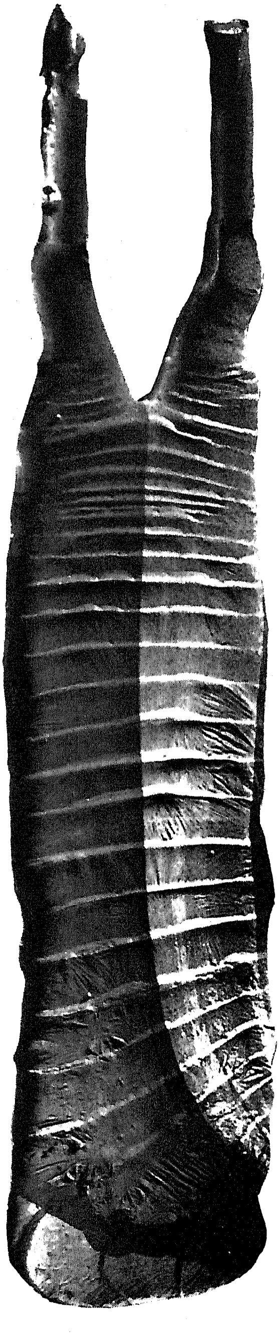


FIGURE 2

observations on the 'macro-mosaic' in these metals are recorded. In addition, the possibility of growing crystal boundaries in chosen directions has been used as the basis for a preliminary investigation of the relative energies of boundaries between crystals with different orientations and of the movement of crystal boundaries under conditions in which the reduction of area of the boundary is the controlling factor.

2. TECHNIQUE

The techniques used in the present work are a development of those described in the paper referred to above. The metal (tin of 99·986 % purity or lead of 99·999 % purity) is melted in a horizonal boat and cooled from one end, at which a seed crystal has previously been inserted. It is necessary that fusion of part of the seed crystal should take place; it is convenient to attach cooling fins to the seed crystal in order that it should not be melted completely. If a boundary between two crystals is required, two seeds of the appropriate orientations are used. The arrangement is illustrated diagrammatically in figure 1.

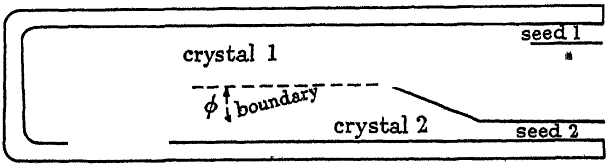


FIGURE 1

Heating is by a coil of a few turns of nichrome strip loosely surrounding the boat. The coil is moved longitudinally and the boat remains stationary. The rate at which the coil is moved can be controlled, and this determines the rate at which the crystals are formed. The rate of growth of the crystals may be observed from time marks produced at appropriate intervals by agitating the surface of the liquid metal. This causes a ripple to be produced which is frozen in at the instantaneous position of the solid-liquid interface. The direction and position of the interface can be observed subsequently from these marks.

The conditions for the growth of single crystals were not critical, and no special temperature control or voltage stabilization was required; single crystals were grown at rates as great as 2 cm./min., and the only condition appears to be that the solid-liquid interface should remain reasonably flat.

The direction of the isothermals can be controlled by suitable positioning of the heating coil, which may be placed alongside instead of round the boat. The effect of doing so is shown in figure 2, plate 1, from which it is seen that the crystal boundary makes a constant angle with the isothermal time marks. In this example, the time marks were at half-minute intervals.

For some of the experiments it was desired to produce a crystal of one orientation between two parts of a second crystal of another orientation. The procedure in this case is to start with a single seed crystal which is grown round an island in the boat. The second seed crystal is introduced where the two parts of the original crystal rejoin. This technique is referred to as 'central seeding', and is illustrated in figure 3.

A simple extension of the same technique allows three crystals to be grown side by side.

The crystal boundaries formed by these methods are always approximately perpendicular to the free surface of the metal.

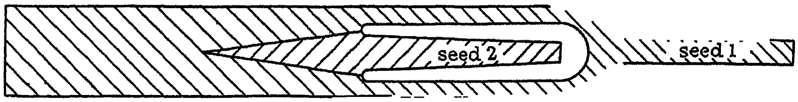


FIGURE 3

3. EXPERIMENTAL RESULTS AND DISCUSSION

Since four distinct phenomena have been studied, it is convenient to describe and discuss the experimental results under the following headings: (*a*) direction of boundary formation, (*b*) macro-mosaic, (*c*) boundary migration, (*d*) angles between boundaries at their junctions.

Direction of boundary formation

It had been observed previously that it is only in special cases that the boundary between two crystals is formed in a plane that is perpendicular to the isothermals, i.e. parallel to the direction of growth of the crystals. This occurs when the two crystals are so related that the axes of one may be made to coincide with those of the other by rotation about the specimen axis, i.e. when a plane perpendicular to the specimen axis has the same crystallographic indices when referred to each of the two crystals. When the relationship is such that the two crystals are not similarly oriented with respect to the specimen axis, the boundary is formed at an angle, denoted by ϕ, to the axis of the specimen (see figure 1). The variation of ϕ with the relative orientations of the two crystals has been examined for tin in the case in which one of the crystals has its c-axis perpendicular to the specimen axis in the plane of the specimen. The orientation of the second crystal is that which would be obtained by rotation of the first crystal through an angle θ about a line in the plane of the specimen perpendicular to the specimen axis. For values of θ other than zero, ϕ, as shown in figure 1, is positive if the c-axis of crystal 1 (figure 1) is perpendicular to the plane of the diagram, and the c-axis of crystal 2 is inclined to this direction by an angle θ. It was found that ϕ increases with θ, the maximum value of ϕ being about 45°, corresponding to $\theta = 90°$. The detailed form of this relationship has not been investigated. Qualitatively similar conclusions have been reached with lead, although no quantitative results have been obtained.

The explanation tentatively advanced when the phenomenon of inclined boundary growth was first noticed (Chalmers 1940) was in terms of the differing thermal conductivities of the two crystals in the relevant direction. It has now been shown that this hypothesis is untenable both for tin and for lead by the following experiments; since the crystal lattice of tin has fourfold symmetry about the c-axis the thermal conductivity is equal in all directions in planes containing the a and b directions. Accordingly, bicrystals were grown in which the c-axis was perpendicular to the plane of the specimen for both crystals while the *a*- and *b*-axes were parallel and perpendicular to the specimen axis in one crystal and inclined at 45° in the other. The boundary was found to slope into the crystal which had its *a*- and *b*-axes at 45° to the specimen axis. It is therefore clear that a difference in thermal conductivity cannot be a necessary condition for inclined boundary growth. In the case of lead, which has a cubic structure, there cannot be any difference of thermal conductivity in different directions; the proposed explanation cannot, therefore, be supported. Further experiments were made to determine whether the solid-liquid interface was continuous from one crystal to the other. The interface that exists at a given moment can be preserved for detailed examination by rapidly tilting the 'boat' so that the liquid metal runs away from the interface. Such interfaces were examined, and it was found that there is always a 'step' where an inclined boundary meets the interface. The direction of the 'step' is shown diagrammatically in figure 4. An estimate has been made of the height of the step in the case of tin; in the extreme case ($\theta = 90°$) the height is about 0·01 mm., while in the case in which both crystals have their *c*-axes perpendicular to the specimen, and the *a*- and *b*-axes vary by 45°, the height is of the order of 0·001 mm. The height of the 'step' was estimated from measurement of the width of the shadow cast by the 'step' under extremely oblique illumination.

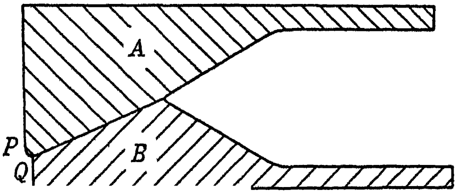


FIGURE 4

The existence of such a step is regarded as accounting for the inclined growth of the boundary, since any accretion on the face *PQ* of figure 4 would cause the crystal *A* to encroach on the crystal *B*.

It is therefore necessary to account for the existence of the 'step', which has been shown to occur even when the thermal conductivities of the two crystals are identical in the plane of the specimen. It follows that the whole of the solid-liquid interface cannot be at the same temperature, since the interface is discontinuous as regards slope. The growing surfaces of the two crystals are therefore at different temperatures. This may be due either to different amounts of supercooling being necessary to cause accretion on them at the common rate of growth imposed by external conditions, or it may be due to an actual difference in the temperature at which solid and liquid are in equilibrium. In the former case, there should be no 'step' in the solid-liquid interface while melting is taking place or while the interface is stationary; if the latter hypothesis is true, there should be a 'step' whether melting or freezing is taking place. It was found in numerous experiments that a 'step' is present if the liquid is decanted during melting or when the interface is stationary. It is indistinguishable both in direction and in size from that observed during solidification. It is therefore concluded that there is a difference between the temperatures of equilibrium of the two crystals with the liquid. This may be accounted for either as a variation of melting-point as a function of the crystallographic orientation of the exposed face, or as a difference in the inherent melting-points of the two crystals. The latter possibility was suggested by Mott (1948), who also suggested the following experiment as a method of distinguishing between the two possibilities. Since the 'step' appears during melting as well as during freezing, it is possible for melting to take place with the isothermal surface inclined to the direction it had during solidification. If the 'step' is a result of a difference between the melting-points of the crystals as such, then it should not be affected by an alteration in the direction of the isothermal; on the other hand, if the crystallographic characteristics of the solid-liquid interface are the controlling factor, then the 'step' should change its direction if the direction of the interface is suitably altered. Experiments on these lines showed conclusively that the direction of the 'step' formed during melting depends on the relative crystal orientations of the interfaces during melting and not by the conditions under which the crystals were formed.

It follows that the temperature of equilibrium between a crystal and the liquid varies with the crystallographic orientation of the interface, and may be different for different faces of the same crystal.

Two phenomena of interest in connexion with the solidification of metals may be accounted for in terms of this effect. In the first place, it has been shown by Northcott (1937, 1938; Northcott & Thomas 1939) that preferred orientations exist in the columnar zone of ingots of copper and of zinc. The columnar zone is a region in which the crystals grow with an advancing solid-liquid interface. It is to be expected that the inclination of the boundary between crystals of different orientations would cause the less favourably oriented crystals to grow for a short distance only before being suppressed by the inclined growth of their more favourably oriented neighbours.

Secondly, the process of dendritic growth of crystals is difficult to account for unless a mechanism is introduced which explains the tendency of the crystal to grow in specific directions rather than in the direction of movement of the general isothermal surfaces. The condition for dendritic growth would then be that the temperature gradient is less than a critical value; this value is such that a 'growing point' with a surface having a raised melting-point would influence the local directions of the isothermals; the greater thermal conductivity of the solid than of the liquid is also an essential factor, since the 'growing point' projecting into the liquid must be cooler than the liquid surrounding it. If the temperature gradient is too great, a 'growing point' would rapidly reach a region where its temperature would exceed its melting-point, where it would be suppressed.

It is possible to estimate the magnitude of the difference of the melting-points of two crystal faces from the temperature gradient in the metal and the height of the step. In a typical example with tin the temperature gradient near the solid-liquid interface was about 1° C/mm. Combining this figure with the estimated height of the step of 0·001 mm., it was found that the temperature difference is of the order of 0·001° C. It should therefore be possible, under sufficiently precise temperature control, to maintain a crystal in the liquid metal so that some faces are freezing while others are melting. The numerical estimates suggest, however, that these conditions would be difficult to obtain.

If an attempt is made to interpret these results in crystallographic terms, it seems that the interface temperature is higher for closer packing of atoms on the exposed face.

Thermodynamically, the melting-point may be looked upon as the temperature at which the free energy per atom in the liquid is equal to that of the solid with which it is in contact. It follows that the free energy per atom at the surface of a tin crystal is a function of the crystallographic characteristics of the face concerned. A corollary is that the latent heat of fusion also depends on the crystallographic characteristics of the face concerned, since the energy change required when an atom originally in the interior of the solid is converted to liquid at a given temperature must be independent of the crystal face through which it passes. Since the specific heats of solid and liquid are not the same, a difference in latent heat must also exist.

As the melting-point may be regarded as the temperature at which the free-energy curves for solid and liquid intersect, it may also be considered that, at temperatures below the melting-point, there will at any moment be some atoms with energies characteristic of the liquid state, while some atoms in the liquid will, even above the melting-point, have energies corresponding to the solid state.

It is considered, therefore, that two competing processes occur at temperatures near the melting-point, namely, the escape of atoms from the solid lattice and the capture of atoms by the lattice. Both these processes may be presumed to occur both above and below the melting-point, which is the temperature at which the two processes occur at equal rates. Atoms which escape from the solid below this temperature are free to move about on the surface or to evaporate, while above the melting-point the reverse tendency results in the formation of temporary clusters or aggregates, which are responsible for short-range order phenomena. The rate of each of the two processes will be controlled by an activation energy; and since the rates are equal at only one temperature, the temperature variations, and therefore the activation energies of the two processes, must be different. The 'melting' process, which predominates at higher temperatures, must vary more rapidly with temperature and so must have a higher activation energy. This consideration shows that the two processes are physically distinct, and it may be inferred that the 'melting' process involves a higher degree of 'co-operation' than the 'freezing' process. It follows that the geometrical relationships of the neighbouring atoms are of more significance in determining the probability of escape of atoms from the solid than for the inverse process; it is therefore suggested that it is the melting process rather than the freezing process which depends upon the crystallographic characteristics of the solid face.

If these considerations are related to the experimental results quoted previously, it may be inferred that the energy required for an atom to escape from the lattice is greater for more closely packed faces than for those with more open packing.

The experiments of Andrade & Martindale (1935) on the crystallization of films of gold and silver suggest an explanation of the difference in freezing-points. They found that (111) faces tend to develop in preference to other faces; that is to say, it is more difficult to add atoms to the (111) face than to other faces. It is to be expected, therefore, that a lower temperature would be necessary to cause liquid atoms to freeze on to such a surface than on to other surfaces. Andrade & Martindale proposed an explanation of this phenomenon, based on Kossel's (1927, 1928) treatment of an idealized case of a simple cubic lattice. The theory is based on considerations of the energy changes that occur when an atom is added to a crystal. The total change of energy is fixed, but the portion of this energy which is involved if the atom is required to start a new plane of atoms depends upon which plane is to be started. From a simple analysis based on the number and distance of the neighbouring atoms, it was shown that more energy is required for an atom to start a new (111) plane than a new (100) or (110) plane.

The corresponding analysis in the case of tin would be more complex, and it is doubtful whether the simplification inherent in the theory of Andrade & Martindale would be permissible for the much less symmetrical tin lattice; but the theory does suggest a physical reason for a variation of melting-point with crystallographic characteristics.

It has been demonstrated previously (Chalmers 1940) that, in the case of tin, the boundary has a melting-point lower by 0·14° C than the crystal's. It follows that, in addition to the 'step' at the solid-liquid interface, there must also be a groove of a depth of about 0·14 mm., but of a width equal to the thickness of the boundary. If, as is generally supposed, this thickness is not more than a few atomic diameters, such a groove would not be observable.

Macro-mosaic

The 'macro-mosaic' effect has not been fully investigated, but the following observations, which were made on specimens prepared for the other experiments described in this paper, throw some light on the structures observed. Most of the crystals both of tin and of lead reveal bands approximately 1 mm. wide running along the length of the specimen; the orientations of the bands are such that one may be transformed to another by a rotation of axes about the specimen axis by about 1°. The structure of the 'single crystal' is therefore a fibre structure, since the component parts have a common axis. The depth of these regions of fixed orientation is about equal to their width, and the whole bulk of the material is made up of 'bars' of these lateral dimensions generally running the whole length of the specimen. It has further been established that the boundaries between the bands behave in a manner generally similar to ordinary crystal boundaries, in so far as their direction is controlled by thermal and not by crystallographic conditions. The boundaries between the bands are, in nearly all cases, perpendicular to the isothermals, as would be expected from their relative orientations.

The origin of this effect has not been explained, but it may be significant that the structure does not appear in the region within about a centimetre of a sloping boundary, in the part of the crystal that grows from the boundary along the direction of growth. They begin to appear after about a centimetre of crystal has formed. The spacing and other characteristics are not influenced by the rate of advance of the solid-liquid interface.

Boundary migration

It is well known that crystal boundaries move during the process of grain growth. The conditions necessary for this to occur are that the metal must have been previously deformed, and it is reasonable to believe that the boundary moves so that a crystal with a lower free energy grows at the expense of one that, as a result of plastic strain, has higher free energy. The generally accepted fact that no boundary movement occurs in 'as cast' structures supports this view.

If, however, the boundary itself possesses free energy, it should tend towards a minimum area. In order to test this hypothesis, experiments were carried out in which the boundary could decrease in area as a result of migration, without any other energy changes taking place. Since tin is not isotropic with respect to thermal expansion, it is necessary to arrange the specimen so that no stresses are introduced during heating or cooling, as such stresses could produce plastic strains which would vitiate the results. Accordingly, specimens were prepared with the c-axes of the two crystals similarly oriented with respect to the axis and plane of the specimen. The 'central seeding' technique was used, the central crystal having its a- and b-axes at 45° to the specimen axis while the outer crystal had these axes along and perpendicular to the specimen axis. A specimen of this type is shown diagrammatically in figure 3. When such a specimen is held at a sufficiently high temperature, movement of the boundary occurs in the expected direction, namely, that which results in a reduction of the area of the boundary. The boundary moves into the inner crystal, the curvature of the tip remaining reasonably uniform. The rate of movement of the boundary can be determined from the positions of the tip after various times. It is found that movement proceeds regularly for considerable periods, but that it eventually stops; although the geometry of the system has apparently not changed, movement may restart spontaneously if the specimen is held at the appropriate temperature for sufficient time.

It is to be expected that the abnormal lattice spacings that occur at crystal boundaries will allow foreign atoms to fit in there more easily than in the lattice of the crystals, especially if the difference in atomic diameter is large. Thus it is probable that the concentration of such atoms will be greater at the boundaries than elsewhere. It is unlikely that the boundary will be able to migrate faster than the atoms associated with it.

It is possible that the observed slowing down of boundary movement is due to the accumulation of an abnormally high concentration of impurity atoms from the region swept by the boundary; further movement may be delayed until the concentration has reached its equilibrium value by diffusion of the excess atoms.

Equilibrium angles between boundaries

In the case in which a boundary can move in such a way that its area is reduced, the activation energy of the process by which the movement occurs could in principle be calculated from experimental data. The actual energy per unit area of the boundary cannot, however, be deduced.

It is possible, nevertheless, to examine the relative energies of different boundaries by measuring the equilibrium angles between three boundaries at the line where they meet. It is of interest to determine whether the energy of a boundary depends on the relative orientation of the two crystals between which it is formed.

By virtue of the fact that the direction of formation of a boundary depends on the relative orientations of the two crystals, it is possible to produce boundaries between three crystals of predetermined orientations, and to cause the three boundaries to meet. Numerous experiments on tin were carried out in this way; in some the relation between the axes of crystals P and Q (figure 5) was that of a fibre structure obtained by rotating the c-axis through small angles (minimum value about 2°) about the specimen axis. In others the rotation was about two axes. These examples represent the case where the PQ boundary had a much smaller difference of orientation than the PR and QR boundaries. In other cases the differences of orientation across the three boundaries were more nearly equal. The specimens were examined microscopically, and the results may be considered under two headings: (*a*) as prepared, and (*b*) after heating.

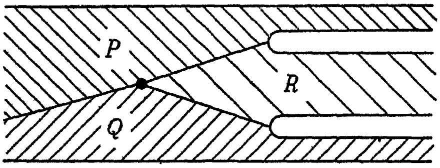


FIGURE 5

(*a*) The angles between the three boundaries were always between 105° and 135° as far as could be judged, in view of the fact that the boundaries near the common point were often curved. This curvature can be regarded as evidence that some approach to equilibrium had taken place after the formation of the boundaries, as they should be straight when formed, and at angles determined by the directions of boundary formation and not by their relative energies.

(*b*) It is known from the work described in the previous section that the boundaries have considerable mobility at temperatures near the melting-point. Specimens were therefore subjected to a temperature about 10° C below the melting-point for periods of 24 hr. and then re-examined. In all cases the angles between the boundaries were within one degree of 120°. This is interpreted to mean that the minimum energy condition is satisfied when the three angles are equal; that is to say, the specific energies of the three boundaries are in all cases equal.

There is some evidence that this result does not apply to the case of a twin boundary meeting an ordinary boundary. It has not been possible so far to produce such specimens under closely controlled conditions, but several specimens have been examined in which the necessary conditions have arisen by chance. It is apparent from such specimens that the twin boundary has a finite specific energy which is probably of the order of one-tenth of that of the ordinary boundary.

Although it has been concluded that all boundaries in tin, except twin boundaries, have the same specific free energy, there are, nevertheless, differences in properties between different boundaries, and these differences appear to be related to the relative orientations of the crystals. It has been observed in many of the specimens that some boundaries etch very much more deeply than others, although no general relationship has been found between this property and the orientation; similar observations have been reported in connexion with other metals (Forsyth, King, Metcalfe & Chalmers 1946), from which it was concluded that the amount of precipitation and the extent of thermal etching are influenced by the relative orientations of the two crystals and the boundary.

This work was carried out at the Atomic Energy Research Establishment of the Ministry of Supply, to whom the author is indebted for permission to publish. The author wishes also to express his indebtedness to Professor E. N. da C. Andrade, F.R.S., for his continued interest in and helpful advice about this research.

REFERENCES

Andrade, E. N. da C. & Martindale, J. G. 1935 *Phil. Trans.* A, **235**, 69.
Chalmers, B. 1940 *Proc. Roy. Soc.* A, **175**, 100.
Forsyth, P. J. E., King, R., Metcalfe, G. J. & Chalmers, B. 1946 *Nature*, **158**, 875.
Kossel, W. 1927 *Mech. Ges. Wiss. Gott*, p. 135.
Kossel, W. 1928 *Phys. Z.* **29**, 553.
Mott, N. F. 1948 *Private communication.*
Northcott, L. 1937 *J. Inst. Met.* **60**, 229.
Northcott, L. 1938 *J. Inst. Met.* **62**, 101.
Northcott, L. & Thomas, D. E. 1939 *J. Inst. Met.* **65**, 205.

A general kinetic theory of liquids
VI. The equation of state

By A. E. Rodriguez, *Edinburgh University*

(*Communicated by M. Born, F.R.S.—Received* 28 *June* 1948)

The present paper is concerned with the study of an approximate equation of state, first derived by Green, which covers both phases, liquid and gas, and can be used for numerical evaluation. It contains explicitly the roots of a certain transcendental equation. The difference between the liquid and the gas corresponds to the existence or non-existence of real roots of this equation.

The solution of the transcendental equation has been reduced, by using a suitable expansion, to the solution of an algebraic equation. In this way explicit expressions for the equation of state are obtained which, however, are too involved to be generally discussed.

The theory can be applied to argon using the Lennard-Jones potential for the interaction between argon atoms, and the results are shown in diagrams. The character of the singularity separating liquid and gas can be seen from these diagrams to lie at the lowest point of the isotherm in the unstable region. An estimate of the position of the critical point is made and found in fair agreement with the experimental data.

INTRODUCTION

In the first parts of this series (Born & Green 1946; Green 1947; Born & Green 1947*a*, *b*) a satisfactorily general kinetic theory of liquids capable of describing the dynamical as well as the equilibrium properties of a statistical assembly of molecules has been developed. In the second of these papers (Green 1947) some attention has been devoted to the problem of condensation, and a new equation of state was obtained, valid for both phases, liquid and gas.

The present work is particularly concerned with the study of that equation of state, with a view to adapting it for numerical evaluation. In the course of this investigation it has been found that some thermodynamical expressions, including that of the equation of state, as given in the original paper (Green 1947) are not quite correct. A new equation, which keeps the main features of the old one, has now been obtained, and is given in § 3.

The final expressions, giving the internal and the free energy, and the equation of state itself, involved the solution of some integro-differential equations, in order to solve which several assumptions were made. It has been considered important to include a section discussing the extent to which these assumptions may influence the final range of validity of the thermodynamical expressions. Most of these results are contained in the paper by Green, mentioned above.

The equation of state here obtained is, of course, approximate; but it covers both phases, liquid and gas, this being the main feature of the theory. The difference between the liquid and the gas corresponds to the existence or non-existence of real roots of a certain transcendental equation. This equation, which in the original form contains an integral in the left-hand side, has now been reduced, by using a suitable expansion, to an algebraic, rapidly convergent series. In this form it is easy to formulate its analytic continuation without falling into ambiguities of any kind. The expansion also provides the way of finding the complex and real roots of the transcendental equation.

The theory of condensation developed by Mayer (1937) and others (Born & Fuchs 1938; Kahn 1938) is confirmed by the present theory to the extent that a divergence of the cluster series is found, which is intimately associated with the process of condensation. On the exact point at which this divergence occurs detailed calculations do not confirm Mayer's contention that it happens at the density of the saturated vapour. The isotherm is continued into the metastable region above this density, rising to a maximum and falling to a subsequent minimum as the density is increased, much in the way foreshadowed by van der Waals over fifty years ago. The point of divergence is found to separate the metastable states of the liquid from those of the gas. This point, which from our standpoint may be considered as a sort of branch point, is identified with that at which one of the complex roots of the transcendental equation becomes real. In this way the thermodynamical properties are regulated through the nature of the roots of this equation.

The main object of the present paper is connected with the solution of this transcendental equation. As has been said, the author has succeeded in reducing the problem to the solution of an algebraic equation which is of $2n+1$th degree, $2n+1$ being the number of terms used in the approximation. One of the sections is entirely devoted to a detailed exposition of the method in the most general form.

In § 6, the equation of state is given in terms of an expansion, to show how the method permits the actual evaluation of some integrals involving the complex roots of the transcendental equation.

Finally a numerical computation has been carried out for argon; two isotherms have been calculated, one for the critical temperature, the other for a somewhat lower one.

Apart from the assumptions which are discussed in the first section, there has been made a rather drastic approximation, namely only the first three roots of the transcendental equation have been taken into account. Better results could perhaps be obtained by taking into account as many as five roots. But in spite of the rather crude numerical computation actually employed, it is clearly shown that the approximation is able to penetrate beyond the singularity into the liquid region.

The expression for the free energy given by Green (1947), which incidentally is not quite correct, can also be expressed with the help of the approximation developed here. One deviation from Green's original method must be mentioned here. He simplified the expression by replacing under some integrals the radial distribution function by its mean value and assumed this to be constant. This approximation turned out to be too crude. Here it will be improved by regarding the mean value as a function of temperature and density. For the perfect gas and effectively above the Boyle point, at which temperature the first virial coefficient vanishes, this parameter becomes identical with unity; but for lower temperatures it may vary quite widely. It is a significant fact that it is by the introduction of this dependence on temperature and density of the parameter, that the critical volume as obtained from our calculations is nearly equal to that given by experiment. If, on the contrary, the value of the parameter were assumed to be constant and equal to unity for all values of temperature and density, the critical volume would turn out to be twice as large as the actual experimental value.

Further, it can be inferred from experiment that the branch point should move with decreasing temperature to higher densities; this would not result from the theory if the parameter is taken as a constant, but is obtained correctly if it is determined as a function of temperature as described.

1. THE FUNDAMENTAL EQUATION. THE KIRKWOOD APPROXIMATION

It has been found by Born & Green (1946) that the radial distribution function, denoted by n_2/n_1^2, satisfies in thermodynamical equilibrium, the equation

$$\frac{\partial n_2(\mathbf{x}^{(1)},\mathbf{x}^{(2)})}{\partial \mathbf{x}^{(2)}}+\frac{n_2(\mathbf{x}^{(1)},\mathbf{x}^{(2)})}{kT}\,\frac{\partial \phi^{(1,2)}}{\partial \mathbf{x}^{(2)}}+\int\frac{n_3(\mathbf{x}^{(1)},\mathbf{x}^{(2)},\mathbf{x}^{(3)})}{kT}\,\frac{\partial \phi^{(2,3)}}{\partial \mathbf{x}^{(3)}}\,d\mathbf{x}^{(3)}=0, \tag{1·1}$$

where $n_3(\mathbf{x}^{(1)},\mathbf{x}^{(2)},\mathbf{x}^{(3)})\,d\mathbf{x}^{(1)}d\mathbf{x}^{(2)}d\mathbf{x}^{(3)}$ is the probability that three distinct molecules 1, 2, 3 occupy volume elements $d\mathbf{x}^{(1)}$, $d\mathbf{x}^{(2)}$, $d\mathbf{x}^{(3)}$, located at $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, $\mathbf{x}^{(3)}$ respectively.

In order to solve (1·1) the following assumption, due to Kirkwood & Boggs (1942), is made about n_3:

$$n_3(\mathbf{x}^{(1)},\mathbf{x}^{(2)},\mathbf{x}^{(3)})=\frac{n_2(\mathbf{x}^{(2)},\mathbf{x}^{(3)})\,n_2(\mathbf{x}^{(3)},\mathbf{x}^{(1)})\,n_2(\mathbf{x}^{(1)},\mathbf{x}^{(2)})}{n_1(\mathbf{x}^{(1)})\,n_1(\mathbf{x}^{(2)})\,n_1(\mathbf{x}^{(3)})}, \tag{1·2}$$

which expresses that the relative probability $n_3/n_1^{(3)}n_2^{(1,2)}$ of the occurrence of a molecule 3 in conjunction with two others should be the product of the relative probabilities $n_2^{(3,1)}/n_1^{(3)}n_1^{(1)}$ and $n_2^{(3,2)}/n_1^{(3)}n_1^{(2)}$ of the occurrence of the same molecule in conjunction with each of the others separately.

Equation (1·2) holds exactly in the case of dilute gases and although it is only approximate in the case of liquids, there are some reasons to presume that one can rely on its validity even for high densities and low temperatures. It is known to have led Kirkwood & Boggs (1942) to results in accordance with experimental data provided by the study of X-ray scattering by monoatomic liquids. It can further be shown* to hold rigorously for solids in the following sense: assuming that the particles have equilibrium positions forming a perfect lattice and neglecting terms in the potential energy higher than the second order, then the probability for finding any three particles with given displacements is expressed by the Kirkwood formula (1·2), in terms of the probabilities for displacements of pairs.

Introducing (1·2) in (1·1) one obtains

$$\frac{\partial n_2(\mathbf{x}^{(1)},\mathbf{x}^{(2)})}{\partial \mathbf{x}^{(2)}}+\frac{n_2(\mathbf{x}^{(1)},\mathbf{x}^{(2)})}{kT}\frac{\partial \phi^{(1,2)}}{\partial \mathbf{x}^{(2)}}=\frac{n_2(\mathbf{x}^{(1)},\mathbf{x}^{(2)})}{n_1^3 kT}\int n_2(\mathbf{x}^{(2)},\mathbf{x}^{(3)})\,n_2(\mathbf{x}^{(3)},\mathbf{x}^{(1)})\frac{\partial \phi^{(2,3)}}{\partial \mathbf{x}^{(3)}}\,d\mathbf{x}^{(3)}, \tag{1·3}$$

(1·3) can be finally brought to the form (see appendix II in paper by Green 1947),

$$\log\frac{n_2(r)}{n_1^2}+\frac{\phi(r)}{kT}=\frac{\pi}{n_1^3}\int_0^\infty\int_{-s}^{s}(s^2-t^2)\frac{t+r}{r}\{n_2(t+r)-n_1^2\}\,dt\,n_2(s)\frac{\phi'(s)}{kT}\,ds. \tag{1·4}$$

The two functions $n_2(r)$ and $\phi(r)$ depend only on the distance r between the two molecules and are defined for negative values to be even functions of r.

2. THE SOLUTION OF THE INTEGRAL EQUATION

The solution of equation (1·4) requires still a new approximation. A glance at (1·4) shows immediately that the integrand contains $n_2(r)$ quadratically. In order to linearize the equation a procedure is used which differs a little from that of Green (1947); it is assumed that $n_2(r)$ is of the form

$$n_2(r) = n_1^2\,\mathrm{e}^{-\phi(r)/kT+f(r)}. \tag{2·1}$$

Inserting (2·1) in (1·4) one obtains

$$\begin{aligned} rf(r) = &-\pi n_1\int_0^\infty\int_{-s}^{s}(s^2-t^2)\,(t+r)\,f(t+r)\,\{1+\alpha(t+r)\}\,dt\,\alpha'(s)\,\{1+f(s)\}\,ds \\ &-\pi n_1\int_0^\infty\int_{-}^{s}(s^2-t^2)\,(t+r)\,\alpha(t+r)\,dt\,\alpha'(s)\,\{1+f(s)\}\,ds, \end{aligned} \tag{2·2}$$

where

$$\alpha(r) = \mathrm{e}^{-\phi(r)/kT}-1. \tag{2·3}$$

The left-hand side of (2·2) follows rigorously, the right-hand side is only approximate, since cubes and higher powers of $f(r)$ have been neglected. This seems to be very reasonable. In fact, it may be inferred from experimental data that, even for rather low temperatures and high densities, the maximum absolute value of $f(r)$ is of the order of 0·4, which implies that one is here neglecting terms that amount to 1% of the total value.

---

* The proof is due to A. A. Sabry of the Department of Mathematical Physics of Edinburgh University, whom I have to thank for kindly informing me about it.

It will be seen later on, that the final expression for $n_2(r)$ obtained using this approximation is substantially equivalent to that obtained by Montroll & Mayer (1941) for the gas.

Following Green, the fact is used, that $\alpha(r)$ and $\alpha'(r)$ are both negligible except for small r, so that, when multiplied by these factors, $f(r)$ may be replaced by $(\epsilon-1)$, its average value in the neighbourhood of the origin. It has been shown by Green (1947) that $\epsilon = 1$ when one is in the gas region, i.e. for low densities and high temperatures. But it is also clear that ϵ is actually a function of these two variables. One can, however, find an explicit expression for ϵ in terms of both variables, which shows that its value may deviate quite largely from unity for low temperatures. On the other hand it may be considered practically constant with respect to density. This will make it possible to find an explicit expression for the free energy.

After integration by parts, (2·3) becomes

$$rf(r) = 2\pi n_1 \int_0^\infty \int_{-s}^{s} (t+r)\{f(t+r)+\epsilon\alpha(t+r)\}\,dt\,\epsilon\alpha(s)\,s\,ds. \tag{2·4}$$

Equation (2·4) is solved by a Fourier transformation (Green 1947). If one writes

$$rg(r) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} s\,.f(s)\sin(rs)\,ds \tag{2·5}$$

and

$$r\,.\beta(r) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} s\,.\alpha(s)\sin(rs)\,ds, \tag{2·6}$$

then (2·4) becomes

$$\lambda g(r) = \{g(r)+\epsilon\beta(r)\}\,\epsilon\beta(r), \tag{2·7}$$

where

$$\lambda = \frac{1}{(2\pi)^{\frac{3}{2}}\,n_1}. \tag{2·8}$$

We obtain finally

$$r\,.f(r) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} \frac{\epsilon^2 s\beta^2(s)\sin(rs)\,ds}{\lambda-\epsilon\beta(s)}. \tag{2·9}$$

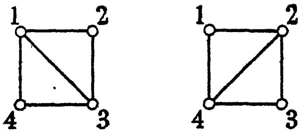
At this point the results may be compared with those of Montroll & Mayer (1941). They have found for the radial distribution function in the gas, which we denote by $n_2^{(M)}(r)$, the following exact expression:

$$n_2^{(M)}(r) = n_1^2\,\mathrm{e}^{-\phi(r)/kT}\left\{1+\frac{1}{(2\pi)^{\frac{1}{2}}r}\int_{-\infty}^{\infty}\frac{s\beta^2(s)\sin(rs)\,ds}{\lambda-\beta(s)}+(2\pi)^{\frac{3}{2}-m}\sum_{m=2}^{\infty}\lambda^{-m}H^{(1)}_{2,m}+\dots\right\}, \tag{2·10}$$

where $H^{(1)}_{2,m}$, $H^{(2)}_{2,m}$, etc., represent integrals over chains of molecules with one internal connexion, two internal connexions, etc., respectively. We have for $m = 2$, for instance,

$$H^{(1)}_{2,2}(\mathbf{x}^{(3)},\mathbf{x}^4) = \frac{1}{2!}\iint[\alpha^{(12)}\alpha^{(23)}\alpha^{(34)}\alpha^{(41)}\alpha^{(13)}+\alpha^{(12)}\alpha^{(23)}\alpha^{(34)}\alpha^{(41)}\alpha^{(24)}]\,d\mathbf{x}^{(1)}d\mathbf{x}^{(2)}, \tag{2·11}$$

which correspond to the well-known 'bond figures'.



If one now expands (2·2) one gets

$$n_2(r) = n_1^2 \, \mathrm{e}^{-\phi(r)/kT} \left\{ 1 + f(r) + \frac{f^2(r)}{2!} + \frac{f^3(r)}{3!} + \dots \right\}. \tag{2·12}$$

Inserting (2·9) one obtains

$$n_2(r) = n_1^2 \, \mathrm{e}^{-\phi(r)/kT} \left\{ 1 + \frac{1}{(2\pi)^{\frac{1}{2}} r} \int_{-\infty}^{\infty} \frac{\epsilon^2 s \beta^2(s) \sin(rs)\, ds}{\lambda - \epsilon\beta(s)} + \dots \right\}. \tag{2·13}$$

Comparing (2·13) with (2·10), it is seen that for the special case $\epsilon = 1$ the two expressions are identical except for terms in higher powers of $f(r)$ which seem to be then equivalent to the terms under the complicated integrals $H^{(1)}_{2,m}$, $H^{(2)}_{2,m}$, etc. apart from the factor ϵ. The effect of averaging and introducing ϵ may therefore be considered as a further refinement. In fact one could still perform an expansion in powers of $(\epsilon - 1)$ and both expressions could be considered equivalent if one identifies the remaining terms, coming on the one hand from higher powers of $f(r)$ and on the other hand from the expansion, with those terms of (2·10) which appear under the summation in the form of complicated integrals. It is a significant fact that ϵ, as will be shown in the next section, only becomes equal to unity for temperatures and densities corresponding to the perfect gas, in which case (2·10) and (2·13) become identical with one another.

It seems that Mayer & Montroll have underestimated the range of validity of their equation (2·10), since, according to what follows in the paper mentioned above, it was only used as an approximation in the gas phase whereas the first two terms, at any rate, possess an analytic continuation into the liquid region.

Assume now that there is an analytical function $\beta(z)$ which reduces to $\beta(r)$ for $z = r$ real. (2·9), then, is the most general solution of (2·4) if the path of integration is chosen as any course from negative to positive infinity in the complex z-plane, and in particular may enclose those points z_u for which the following transcendental equation

$$\beta(z) = \lambda/\epsilon \tag{2·14}$$

is satisfied.

$\beta(z)$ will now be defined for complex z. Assume that $s.\alpha(s)$ admits an expansion of the type

$$s.\alpha(s) = \sum_{k=-\infty}^{\infty} c_k \, \mathrm{e}^{is\xi_k}, \tag{2·15}$$

ξ_k being a complex number in the upper half of the complex plane. This way of expanding the function $s.\alpha(s)$ is based on the fact that for all $s > s_0$ we have practically $s.\alpha(s) = 0$, as shown in figure 1.

By a proper choice of the form of the ξ_k, as will be shown clearly in § 5, $s.\alpha(s)$ can be assumed to be represented by the expansion (2·15).

From (2·6), which gives the definition of $\beta(r)$ for r real, one gets, after inserting (2·15),

$$r.\beta(r) = \frac{2}{(2\pi)^{\frac{1}{2}}} \sum_k c_k I_k, \tag{2·16}$$

where

$$I_k = \int_0^{\infty} \mathrm{e}^{is\xi_k} \sin(rs)\, ds. \tag{2·17}$$

Integration by parts gives, remembering that ξ_k is a complex number in the upper half of the complex plane,

$$I_k = \frac{-r}{\xi_k^2 - r^2}, \tag{2·18}$$

which, introduced in (2·16), gives

$$\beta(r) = -\frac{2}{(2\pi)^{\frac{1}{2}}} \sum_k \frac{c_k}{(\xi_k^2 - r^2)}. \tag{2·19}$$

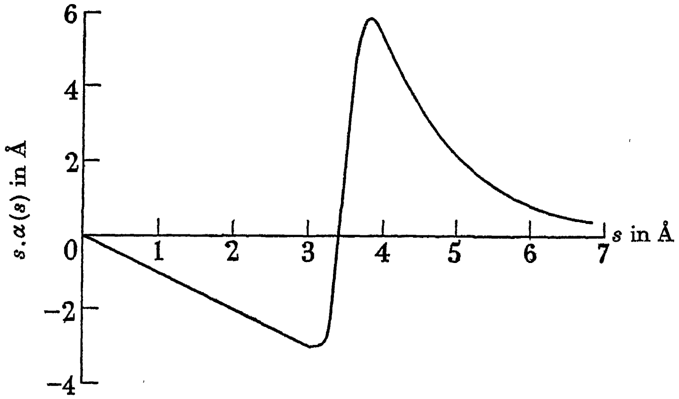


FIGURE 1

The function on the right-hand side is an analytic function of r except for the occurrence of simple poles at the points $\pm\xi_k$. The equation (2·19) may therefore be regarded as a satisfactory definition of the function $\beta(r)$ for all values of r real and complex satisfying the following conditions:

(*a*) $\beta(z)$ reduces to $\beta(r)$ for $z = r$ real,

(*b*) $\beta(z) \to 0$ as $|z| \to \infty$, uniformly with respect to arg z, in the upper half of the z-plane.

3. THE EQUATION OF STATE

The pressure p is given (Green 1947) by

$$p = n_1 kT - \frac{1}{6}\int_0^\infty n_2(r)\,\phi'(r)\,4\pi r^3\,dr. \tag{3·1}$$

In order to obtain an explicit expression for p, one now expands (2·2) and takes only the linear term in $f(r)$, i.e.

$$n_2(r) = n_1^2\,\mathrm{e}^{-\phi(r)/kT}\{1 + f(r)\}. \tag{3·2}$$

This new approximation seems to be reasonable; for $r > r_1$, r_1 being the value of r for which $n_2(r)$ has its absolute maximum, when one substitutes (3·2) in the integral appearing in (3·1), squares and higher powers of $f(r)$ will not give an appreciable contribution, since the effect of the total correction $\mathrm{e}^{f(r)}$ is to superimpose on the first factor $\mathrm{e}^{-\phi(r)/kT}$ oscillations around a constant value, which will cancel one another on integration. Contributions coming from squares and higher powers of

$f(r)$ for $r < r_1$, are of course important, but have been sufficiently taken into account for the present purpose by the introduction of the parameter ϵ.

Writing for $f(r)$ the expression obtained in the last section, namely,

$$rf(r) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\frac{\epsilon^2 s\,.\,\beta^2(s)\sin(rs)\,ds}{\lambda-\epsilon\beta(s)}, \tag{3·3}$$

reversing (2·6),

$$r\alpha(r) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty} s\beta(s)\sin(rs)\,ds, \tag{3·4}$$

and combining (3·3) and (3·4) one gets

$$\begin{aligned} r\{f(r+\epsilon\alpha(r)\} &= \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\left\{\frac{\epsilon^2\beta^2(s)}{\lambda-\epsilon\beta(s)}+\epsilon\beta(s)\right\}s\sin(rs)\,ds \\ &= \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\frac{\epsilon^2\beta^2(s)+\{\lambda-\epsilon\beta(s)\}\,\epsilon\beta(s)}{\lambda-\epsilon\beta(s)}\,s\sin(rs)\,ds \\ &= \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\frac{\lambda\epsilon\,.\,s\,.\,\beta(s)\sin(rs)}{\lambda-\epsilon\beta(s)}\,ds. \end{aligned} \tag{3·5}$$

The form for $f(r)$ as given by equation (3·3) suggests immediately an expansion in powers of the density; in fact for gases where $|\lambda-\epsilon\beta(s)|>0$ for all real s, such an expansion is permissible, and leads to the following result (Green 1947):

$$p = n_1 kT\left\{1-\frac{\beta(0)}{2\lambda}-\sum_{\nu=2}^{\infty}\frac{\nu\epsilon^\nu}{2(2\pi)^{\frac{1}{2}}\lambda^\nu}\int_{-\infty}^{\infty}\frac{\beta^{\nu+1}(s)\,s^2ds}{\nu+1}\right\}. \tag{3·6}$$

An exact development of the equation of state as a power series in the density was made first by Ursell (1927), later improved by Mayer (1937), Born & Fuchs (1938) and given a quantum generalization by Kahn (1938). The result may be written as follows

$$p = n_1 kT\left\{1-\sum_{s=1}^{\infty}\frac{s}{s+1}\beta_s n_1^s\right\}, \tag{3·7}$$

so that the quantities $-\frac{s}{s+1}\beta_s$ are the so-called virial coefficients. The β_s are rather complicated integrals, and so far analytic expressions have been found only for the first three (de Boer 1940).

Mayer (1937) has shown that the equation (3·7) does not describe the liquid phase, simply because the infinite series become divergent on approaching the point of condensation. The same thing happens to (3·6) when the transcendental equation $\beta(z) = \lambda/\epsilon$ first has a real root and so it is clearly seen why the infinite series in (3·7) diverges. As to the exact point where this divergence occurs, however, detailed calculations do not confirm that it happens at the density of the saturated vapour. The point of divergence is found as separating the metastable states of the liquid from those of the gas.

One should conclude, accordingly, that a different method of expansion is required if one wishes to cover the liquid as well as the gas region. For this purpose it is convenient to consider the complex roots z_u in the upper half of the z-plane of the equation $\beta(z) = \lambda/\epsilon$, and use instead of (3·3) the alternative expression (3·5) for $f(r)$.

It follows from the definition (2·19) that the conditions (*a*) $\beta(z)$ reduces to $\beta(r)$ for $z = r$ real, and

(*b*) $\beta(z) \to 0$ as $|z| \to \infty$, uniformly with respect to $\arg z$, in the upper half of the z-plane, are satisfied. By applying a well-known theorem of the calculus of residues, (3·5) becomes

$$rf(r) = -\epsilon . r . \alpha(r) - (2\pi)^{\frac{1}{2}} \frac{\lambda^2}{\epsilon} \sum_u \frac{z_u}{\beta'(z_u)} e^{irz_u}. \tag{3·8}$$

Inserting (3·8) and (3·2) in (3·1) one obtains, after some calculation,

$$p = n_1 kT \left\{ 1 - \frac{\beta(0)}{2\lambda} + \frac{\epsilon}{4(2\pi)^{\frac{1}{2}}\lambda} \int_{-\infty}^{\infty} \beta^2(r)\, r^2\, dr + \frac{2}{3}\frac{\lambda}{\epsilon} \sum_u \frac{z_u}{\beta'(z_u)} \int_0^{\infty} r\alpha(r)\, e^{irz_u}\, dr \right.$$
$$\left. + \frac{i\lambda}{3\epsilon} \sum_u \frac{z_u^2}{\beta'(z_u)} \int_0^{\infty} r^2 \alpha(r)\, e^{irz_u}\, dr \right\}. \tag{3·9}$$

Equation (3·9) holds equally for the liquid and the gas. The difference between them arises from the fact that in the case of the liquid there are two real roots which are not included in the u-summation in (3·9).

In figure 2 the function $\beta(r)$ has been plotted for a certain temperature T below the critical one.

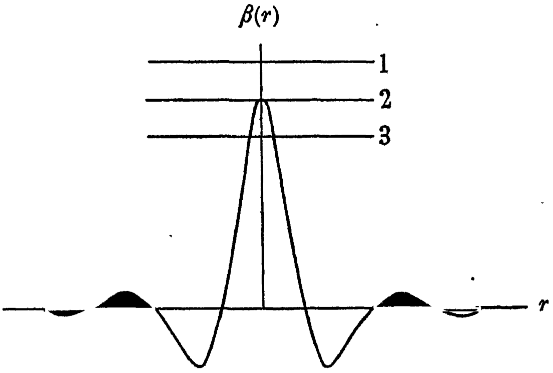


FIGURE 2

For low densities λ/ϵ exceeds the upper bound of $\beta(r)$ (case 1 in figure 2) and the fundamental equation has no real roots. This has to be interpreted as the region of the gas. If the density is increased, while the temperature is kept constant, a certain density is reached (case 2 in figure 2) at which two coincident real roots exist. This point is interpreted as separating the metastable states of the liquid from those of the gas. If the density is increased still more, one has two real roots (case 3 in figure 2) which of course are not included in the u-summation of (3·9). This region for which equation $\beta(z) = \lambda/\epsilon$ has two real roots corresponds to the liquid phase.

For higher temperatures the upper bound of $\beta(r)$, which always occurs at the point $r = 0$, becomes so small that only at exceedingly high densities will the equation $\beta(z) = \lambda/\epsilon$ have real roots and the liquid region can never be reached.

Equation (3·9) is our final equation of state. (It cannot be put in the form given by Green (1947) because

$$\int_0^{\infty} r\alpha(r)\, e^{irz_u}\, dr \neq \frac{1}{2} \int_{-\infty}^{\infty} r\alpha(r)\, e^{irz_u}\, dr \tag{3·10}$$

for all z_u complex in the upper half of the complex plane.)

In some previous considerations the fact has been mentioned that the parameter ϵ, which was introduced to average the function $f(r)$, is not a constant, but a function of temperature and density. This will now be determined.

If the equation (3·8), after multiplication by $r.\alpha(r)$, is integrated over r, one obtains

$$\frac{\int_0^\infty r^2\alpha(r)\,f(r)\,dr}{\int_0^\infty r^2\alpha(r)\,dr}\int_0^\infty r^2\alpha(r)\,dr = -\epsilon\int_0^\infty r^2\alpha^2(r)\,dr-\epsilon(2\pi)^{\frac{1}{2}}\left(\frac{\lambda}{\epsilon}\right)^2\sum_u\frac{z_u}{\beta'(z_u)}\int_0^\infty r\alpha(r)\,e^{irz_u}\,dr. \tag{3·11}$$

The first factor in the left-hand side is by definition equal to $(\epsilon-1)$. Hence

$$\epsilon=\frac{\int_0^\infty r^2\alpha(r)\,dr}{\int_0^\infty r^2\alpha(r)\,dr+\int_0^\infty r^2\alpha^2(r)\,dr+(2\pi)^{\frac{1}{2}}\left(\frac{\lambda}{\epsilon}\right)^2\sum_u\frac{z_u}{\beta'(z_u)}\int_0^\infty r\alpha(r)\,e^{irz_u}\,dr}. \tag{3·12}$$

It is obvious from (3·8) that at high temperatures and low densities $\epsilon \to 1$, since $f(r)$ is actually equal to zero.

For comparatively low temperatures and high densities ϵ will not equal 1, since $f(r)$ may take high values. Detailed calculations made for argon confirm this argument. For this gas at temperatures near the critical temperature $\epsilon = 0{\cdot}60$, and it remains constant practically for all values of density. For the same gas and a temperature $T = 130°\,\mathrm{K}$, $\epsilon = 0{\cdot}46$.

4. THE ALGEBRAIC EQUATION

This section will be concerned with the solution of the fundamental equation

$$\beta(z)=\lambda/\epsilon. \tag{4·1}$$

By using the definition (2·20) given for $\beta(z)$, (4·1) becomes

$$\sum_{k=-n}^{n}\frac{c_k}{\xi_k^2-z^2}=-\frac{(2\pi)^{\frac{1}{2}}}{2}\frac{\lambda}{\epsilon}, \tag{4·2}$$

where the c_k's are the coefficients of the expansion $s.\alpha(s)$ (2·16). The number n will be considered finite. (4·2) is therefore an algebraic equation of order $2n+1$ in z^2.

We write for abbreviation

$$C_k=\frac{2\epsilon c_k}{(2\pi)^{\frac{1}{2}}\lambda}. \tag{4·3}$$

Then (4·2) becomes simply

$$z^{4n+2}-P_1z^{4n}+P_2z^{4n-2}-\ldots\pm P_{2n+1}=0, \tag{4·4}$$

where

$$P_j=\Pi_j+\sum_{k=-n}^{n}C_k\Pi_{j-1}^{(k)}, \tag{4·5}$$

and the following notation has been introduced

$$\left.\begin{aligned}\Pi_1 &= \xi_0^2+\xi_1^2+\xi_{-1}^2+\xi_2^2+\xi_{-2}^2+\ldots+\xi_n^2+\xi_{-n}^2,\\ \Pi_2 &= \xi_0^2\xi_1^2+\xi_0^2\xi_{-1}^2+\ldots+\xi_{-n}^2\xi_n^2,\\ \Pi_3 &= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\\ \Pi_{2n+1} &= \xi_0^2\xi_1^2\xi_{-1}^2\xi_2^2\xi_{-2}^2\cdots\xi_n^2\xi_{-n}^2,\end{aligned}\right\} \tag{4·6}$$

and

$$\left.\begin{aligned}\Pi_0^{(k)} &= 1,\\ \Pi_1^{(k)} &= \Pi_1-\xi_k^2,\\ \Pi_2^{(k)} &= \Pi_2-\xi_k^2\,\Pi_1^{(k)},\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\\ \Pi_j^{(k)} &= \Pi_j-\xi_k^2\,\Pi_{j-1}^{(k)},\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\\ \Pi_{2n}^{(k)} &= \Pi_{2n}-\xi_k^2\,\Pi_{2n-1}^{(k)} = \Pi_{2n+1}/\xi_k^2.\end{aligned}\right\} \tag{4·7}$$

By successive elimination one obtains for the general term (4·7)

$$\Pi_j^{(k)} = \Pi_j-\xi_k^2\,\Pi_{j-1}+\xi_k^4\,\Pi_{j-2}-\xi_k^6\,\Pi_{j-3}+\ldots+\xi_k^{2j}. \tag{4·8}$$

Here the Π's can all be expressed by Π_1 by use of the recurrence formulae

$$\left.\begin{aligned}\Pi_2 &= \frac{1}{2}\left\{\Pi_1\Pi_1-\sum_{k=-n}^{n}\xi_k^4\right\},\\ \Pi_3 &= \frac{1}{3}\left\{\Pi_1\Pi_2-\sum_{k=-n}^{n}\Pi_1^{(k)}\xi_k^4\right\},\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\\ \Pi_j &= \frac{1}{j}\left\{\Pi_1\Pi_{j-1}-\sum_{k=-n}^{n}\Pi_{j-2}^{(k)}\xi_k^4\right\},\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\end{aligned}\right\} \tag{4·9}$$

which can be proved by complete induction with the help of (4·7).

Introducing (4·8) in (4·9) one gets

$$\Pi_j = \frac{1}{j}\left\{\Pi_1\Pi_{j-1}-\Pi_{j-2}\sum_{k=-n}^{n}\xi_k^4+\Pi_{j-3}\sum_{k=-n}^{n}\xi_k^6-\ldots\pm\sum_{k=-n}^{n}\xi_k^{2j}\right\}. \tag{4·10}$$

These recurrence formulae express Π_j by all previous $\Pi_{j-1}, \Pi_{j-2}, \ldots, \Pi_1$.

5. THE EXPANSION OF THE FUNCTION $s.\alpha(s)$

We consider now the expansion (2·15)

$$s.\alpha(s) = \sum_{k=-n}^{n} c_k\,e^{is\xi_k} \tag{5·1}$$

and have to choose the ξ_k's and to determine the c_k's. The ξ_k's have to be complex numbers in the upper half of the complex plane. The simplest choice is an equidistant set of points on a parallel to the real axis

$$\xi_k = \beta.k+i.\gamma \quad (k = -n, \ldots, -1, 0, 1, \ldots, n). \tag{5·2}$$

We then have to choose domain $(0, s_0)$ in which the function $s.\alpha(s)$ is regarded as essentially different from zero while it is neglected for $s > s_0$. The choice of s_0 and γ are related to one another by the following consideration: if one defines

$$\psi(s) = \sum_{k=-n}^{n} c_k \, e^{iks\beta}, \tag{5·3}$$

where $s_0\beta = 2\pi$, one should have, according to (5·1),

$$s.\alpha(s) = e^{-\gamma.s}\,\psi(s). \tag{5·4}$$

$\psi(s)$ is periodic with period s_0 while the $s.\alpha(s)$ is aperiodic. Therefore (5·4) can only approximately hold if $e^{-\gamma.s}$ is negligible for $s > s_0$. Hence we choose $\gamma.s_0 = \frac{2\pi}{\beta}.\gamma$ as a rather big number; but on the other hand $\psi(s_0) = s_0\alpha(s_0)\,e^{\gamma.s_0}$ must be small in order that $\psi(s)$ should be periodic without a big discontinuity, $\psi(0) \sim \psi(s_0)$. Hence one must make a practical compromise for the two numbers β and γ, which can be done only for a given numerical example.

For practical calculations one writes (5·3) in the real form

$$\psi(s) = a_0 + \sum_{k=1}^{n} (a_k \cos s\beta k + b_k \sin s\beta k) \quad (s_0\beta = 2\pi), \tag{5·5}$$

where

$$c_0 = a_0, \quad c_k = \tfrac{1}{2}(a_k - ib_k), \quad c_{-k} = \tfrac{1}{2}(a_k + ib_k). \tag{5·6}$$

From (4·6) and (5·2) one gets

$$\Pi_1 = -(2n+1)\gamma^2 + 2\beta^2 \sum_{k=1}^{n} k^2. \tag{5·7}$$

From (5·2) one obtains further

$$\sum_{k=-n}^{n} \xi_k^{2j} = (-1)^j (2n+1)\gamma^{2j} + \left\{\binom{2j}{2j}\beta^{2j}\sum_{k=1}^{n} k^{2j} - \binom{2j}{2j-2}\beta^{2j-2}\gamma^2 \sum_{k=1}^{n} k^{2j-2} + \binom{2j}{2j-4}\beta^{2j-4}\gamma^4 \sum_{k=1}^{n} k^{2j-4} - \dots \pm \binom{2j}{2}\beta^2\gamma^{2j-2}\sum_{k=1}^{n} k^2\right\}. \tag{5·8}$$

With the help of these formulae one gets all the Π_j's in terms of β, γ and n.

The whole process can be summarized in the following way: the function $\psi(s) = s.\alpha(s)\,e^{\gamma.s}$ is plotted and the constant γ and the interval $(0, s_0)$ chosen such that $\psi(s_0)$ is negligible and still $\gamma.s$ is large enough so that $e^{-\gamma.s}$ times the periodic repetition of $\psi(s)$ can be neglected.

From this s_0, $\beta = 2\pi/s_0$ is found. Then an integer n is chosen depending on the accuracy to be intended. Then Π_1 is calculated from (5·7) and the sums $\sum_{k=-n}^{n} \xi_k^{2j}$ from (5·8). The recurrence formulae (4·10) give $\Pi_2, \Pi_3, \dots, \Pi_{2n+1}$, then the $\Pi_j^{(k)}$ are obtained from (4·7) $(j = 1, 2, \dots, 2n)$.

On the other hand the coefficients a_k, b_k of the function $\psi(s)$ according to (5·5) are determined by one of the known methods of Fourier analysis. Then one has the complex coefficients c_k's from (5·6) and C_k's from (4·3) for a suitable set of values λ/ϵ. Finally the P_j's from (4·5).

Now the algebraic equation (4·4) can be solved numerically and leads to the roots z_u^2. For each value λ/ϵ one has a set of $2n+1$ roots z_u^2 from which the roots in the positive half complex plane, which we need for the further calculation, are selected.

6. The equation of state in terms of the expansion

In the equation of state (3·9) there appear, apart from the z_u, some values of the function $\beta(s)$ and some integrals over the functions $s\,.\,\alpha(s)\,\mathrm{e}^{isz}$ and $s^2\alpha(s)\,\mathrm{e}^{isz}$ taken at the point z_u. One can express all these in terms of the constants c_k's and ξ_k's of the expansion (2·15).

One obtains, from the definition (2·19) for $\beta(s)$,

$$\beta(0) = -\frac{2}{(2\pi)^{\frac{1}{2}}}\sum_{k=-n}^{n}\frac{c_k}{\xi_k^2} \tag{6·1}$$

and

$$\int_{-\infty}^{\infty} s^2\beta^2(s)\,ds = -2\sum_{k,s=-n}^{n}\frac{c_k c_s}{i(\xi_k+\xi_s)}, \tag{6·2}$$

and also from (2·15)

$$\int_{0}^{\infty} s\,.\,\alpha(s)\,\mathrm{e}^{isz_u}\,ds = i\sum_{k=-n}^{n}\frac{c_k}{(\xi_k+z_u)} \tag{6·3}$$

and

$$\int_{0}^{\infty} s^2\,.\,\alpha(s)\,\mathrm{e}^{isz_u}\,ds = \sum_{k=-n}^{n}\frac{c_k}{(\xi_k+z_u)^2}. \tag{6·4}$$

From equation (2·19) for $\beta(z)$ it follows:

$$\frac{\beta'(z_u)}{z_u} = -\frac{4}{(2\pi)^{\frac{1}{2}}}\sum_{k=-n}^{n}\frac{c_k}{(\xi_k^2-z_u^2)^2}. \tag{6·5}$$

As the algebraic equation (4·4) has real coefficients, to every root

$$z_u = s_u + it_u \tag{6·6}$$

belongs another one, the conjugate number, and since the equation contains only powers in z^2, the negative of z_u^* is also a root; now

$$-z_u^* = -s_u + it_u. \tag{6·7}$$

Hence, the roots which are needed, are those of the type (6·6) and (6·7) for which

$$t_u > 0. \tag{6·8}$$

There may be a root on the real axis which may be characterized by the index $u = 0$, so that

$$z_0 = it_0. \tag{6·9}$$

The following abbreviations are introduced:

$$\left.\begin{aligned}
& i\sum_{k=-n}^{n}\frac{c_k}{\xi_k+z_0} = A_0,\\
& \left.\begin{aligned}
i\sum_{k=-n}^{n}\frac{c_k}{\xi_k+z_u} &= A_u+iB_u,\\
i\sum_{k=-n}^{n}\frac{c_k}{\xi_k-z_u^*} &= A_u-iB_u,
\end{aligned}\right\}\quad (u = 1, 2, \ldots, n)
\end{aligned}\right\} \tag{6·10}$$

$$\left.\begin{aligned}&\sum_{k=-n}^{n}\frac{c_k}{(\xi_k+z_0)^2}=A_0',\\&\left.\begin{aligned}&\sum_{k=-n}^{n}\frac{c_k}{(\xi_k+z_u)^2}=A_u'+iB_u',\\&\sum_{k=-n}^{n}\frac{c_k}{(\xi_k-z_u^*)^2}=A_u'-iB_u',\end{aligned}\right\}\quad(u=1,2,\ldots,n)\end{aligned}\right\}\tag{6·11}$$

$$\left.\begin{aligned}&-\frac{4}{(2\pi)^{\frac{1}{2}}}\sum_{k=-n}^{n}\frac{c_k}{(\xi_k^2-z_0^2)^2}=A_0'',\\&\left.\begin{aligned}&-\frac{4}{(2\pi)^{\frac{1}{2}}}\sum_{k=-n}^{n}\frac{c_k}{(\xi_k^2-z_u^2)^2}=A_u''+iB_u'',\\&-\frac{4}{(2\pi)^{\frac{1}{2}}}\sum_{k=-n}^{n}\frac{c_k}{\xi_k^2+z_u^*}=A_u''-iB_u'',\end{aligned}\right\}\quad(u=1,2,\ldots,n).\end{aligned}\right\}\tag{6·12}$$

Finally, equation of state (3·9) becomes, according to (6·1), (6·2), (6·3) and (6·4) and (6·10), (6·11) and (6·12)

$$p=n_1kT\left\{1-\frac{F_1(T)}{(2\pi)^{\frac{1}{2}}\lambda}+\frac{2\lambda}{3\epsilon}F_2(T,\lambda)\right\},\tag{6·13}$$

where

$$F_1(T)=-\sum_{k=-n}^{n}\frac{c_k}{\xi_k^2}+\epsilon\sum_{k,s=-n}^{n}\frac{c_kc_s}{i(\xi_k+\xi_s)}\tag{6·14}$$

and

$$F_2(T,\lambda)=\frac{2A_0-A_0't_0}{2A_0''}+2\sum_{u=1}^{n}\frac{A_u''(2A_u+s_uB_u'-t_uA_u')+B_u''(2B_u-s_uA_u'-t_uB_u')}{A_u''^2+B_u''^2}.\tag{6·15}$$

If one now assumes ϵ to be only a function of temperature (which is practically the case, as numerical evaluation shows), the free energy A is easily obtained from (3·9) by integrating the partial differential equation

$$\frac{\partial}{\partial V}A=-p.\tag{6·16}$$

Using (6·1), (6·2) and (6·3), A can be written in the form

$$A=-\frac{NkT}{2\lambda}\left\{-\frac{2}{(2\pi)^{\frac{1}{2}}}\sum_{k=-n}^{n}\frac{c_k}{\xi_k^2}+\frac{\epsilon}{2(2\pi)^{\frac{1}{2}}}\sum_{k,s=-n}^{n}\frac{c_kc_s}{i(\xi_k+\xi_s)}+\frac{2\lambda}{3}\sum_{k,s=-n}^{n}c_k\left(\frac{z_s^2}{\xi_k+z_s}-\frac{\xi_s^2}{\xi_k+\xi_s}\right)\right\}$$
$$+NkT\log n_1-\tfrac{3}{2}NkT\log T.\tag{6·17}$$

The arbitrary function of temperature which appears on integration is calculated by comparing the formula which one obtains with the well-known formula for the free energy for the case of a very dilute gas (λ large).

Equation (6·17) is of course approximate. It contains not only the approximations which were used in obtaining the pressure p, but it is obtained under the assumption that ϵ is only a function of temperature.

7. APPLICATION. THE EQUATION OF STATE FOR ARGON

The theory will be now applied to argon because here a good representation of the potential energy is available through the work of Lennard-Jones (1924). He has shown that the function

$$\phi(s) = \frac{a}{s^6} - \frac{b}{s^{12}} \tag{7·1}$$

with $$a = 1{\cdot}03 \times 10^{-10}\,\text{ergs}\,\text{Å}^6, \quad b = 1620 \times 10^{-10}\,\text{ergs}\,\text{Å}^{12}, \tag{7·2}$$

allows one to explain the behaviour of the dilute gas as determined by measuring the first virial coefficient. Figure 3 shows a graph of $\phi(s)$.

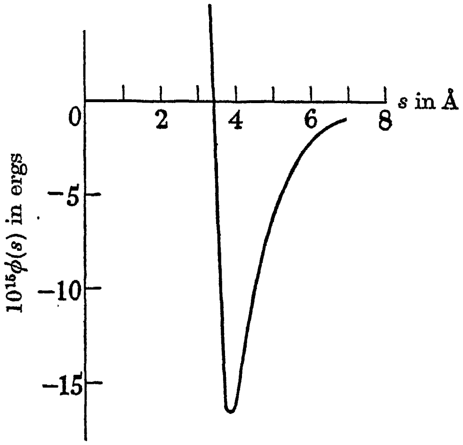


FIGURE 3

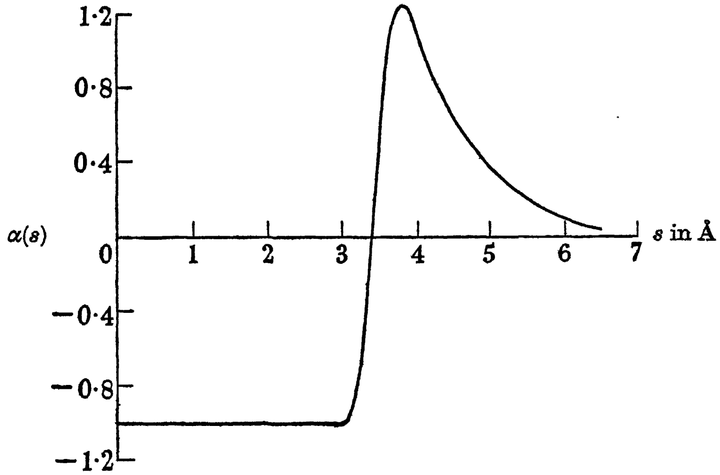


FIGURE 4

The minimum of $\phi(s)$ occurs for a value s_1 of s given by

$$s_1 = \sqrt[6]{\frac{2b}{a}} = 3{\cdot}83\,\text{Å}, \tag{7·3}$$

for which $$\phi(s_1) = -16{\cdot}5 \times 10^{-15}\,\text{ergs}. \tag{7·4}$$

In figure 4 the function $\alpha(s) = e^{-\phi(s)/kT} - 1$ has been plotted for $T = 150°\,\text{K}$.

With the choice $s_0 = 7{\cdot}2\,\text{Å}$, $\alpha(s_0) \simeq 0{\cdot}03$. By trial and error a suitable value for γ has been found to be $\gamma = 0{\cdot}3\,\text{Å}^{-1}$ with which the graph in figure 5 has been drawn. There is still an error in periodicity of $\frac{1}{10}$ relative to the maximum. On the other hand $\gamma s_0 = 2{\cdot}16$ is large enough to make the repetition of the curve for $s > s_0$ negligible as can be seen in figure 6. Hence one obtains

$$\left.\begin{aligned} \beta &= 2\pi/s_0 = 0{\cdot}872\,\text{Å}^{-1}, \\ \gamma &= 0{\cdot}3\,\text{Å}^{-1}, \\ \xi_k &= \beta.k + i\gamma \quad (k = -n, \ldots, -1, 0, 1, \ldots, n). \end{aligned}\right\} \tag{7·5}$$

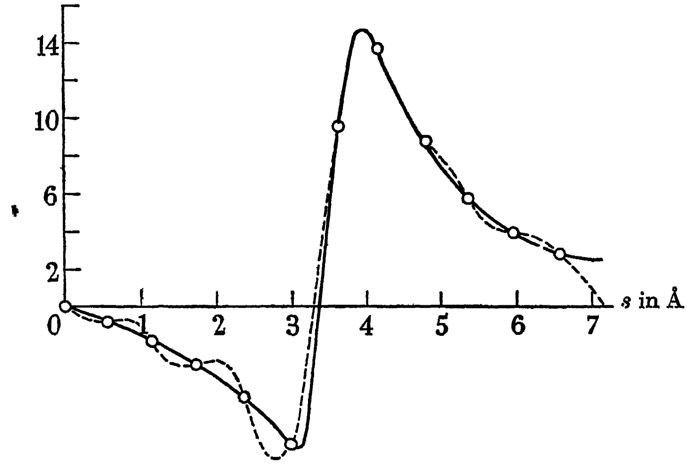


FIGURE 5. —— function $\psi(s) = s.\alpha(s)\,e^{\gamma.s}$; ---- approximation by a Fourier series.

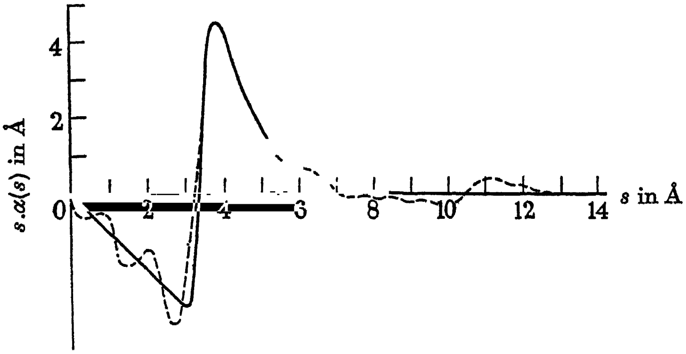


FIGURE 6. —— function $s.\alpha(s)$; ---- approximation.

Now the coefficients of the Fourier expansion (2·16) have been numerically determined (Whittaker & Robinson 1937) by using 12 equally spaced points as shown in the graph of figure 5. The results in Å is given, for $T = 150°\,\text{K}$, in table 1. Figure 5 also gives a comparison of the Fourier series with the actual function $\psi(s) = s\alpha(s)\,e^{\gamma.s}$. From table 1 the c_k's are obtained with the help of (5·6). $\beta(r)$ can then be calculated from (2·20) using all terms up to $k = 6$. The result is given in figure 7. All terms up to $k = 6$ are also used later in calculating the coefficients

A_u, B_u, A'_u, B'_u, A''_u, B''_u according to (6·10), (6·11) and (6·12). With $k = 6$ the equation (4·4) would be of 13th order in z^2, and have 13 roots for z^2 (26 for z). It would be very cumbersome to use them all in the following calculation; it has turned out that the first three roots z_0, z_1, $-z_2^*$ suffice, but these can be determined with sufficient accuracy from a simplified equation where only the terms $k = -1, 0, 1$ in the coefficients are used. This can be seen by examining the figure 7 in which the curve $\beta(r)$ calculated before is compared with that obtained by using only the terms $k = -1, 0, 1$. They coincide practically completely apart from deviations below the abscissa axis, which do not interfere with the roots. This is evident for the real roots, namely the intersection of the curve with the horizontal line as indicated in the figure, but holds also for the complex roots. With this simplification equation (4·4) reduces to

$$z^6 - P_1 z^4 + P_2 z^2 - P_3 = 0, \tag{7·6}$$

where P_j's are obtained by the process described earlier.

TABLE 1

$a_0 = 2·231$	
$a_1 = -2·321$	$b_1 = -6·321$
$a_2 = 1·366$	$b_2 = +3·646$
$a_3 = -1·343$	$b_3 = -2·626$
$a_4 = 0·836$	$b_4 = +1·356$
$a_5 = -1·158$	$b_5 = -0·725$
$a_6 = 0·391$	$b_6 = 0·000$

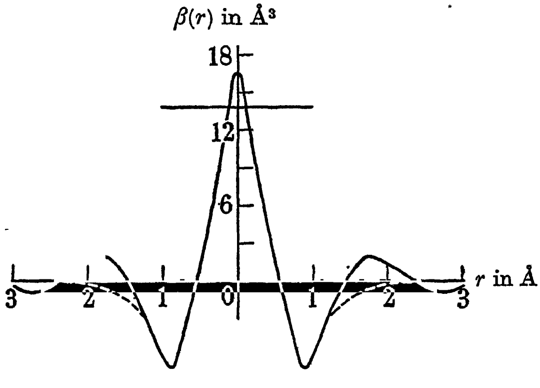


FIGURE 7

Now by choosing a suitable set of values λ/ϵ one obtains, by solving equation (7·6), a corresponding set of values t_0, s_1, t_1 from which all the A_u, B_u, A'_u, B'_u, A''_u, B''_u can be evaluated. In the same way the corresponding values of ϵ are calculated from (3·12) and then the values for λ can be obtained. In table 2 the values of ϵ are given as functions of λ.

Finally, using the equation of state (6·13), the isotherm for $T = 150°$ K has been calculated. In table 3, the pressure p is given as function of volume, for one cubic centimeter of gas at normal pressure and temperature. The pressure p is expressed in atmospheres and the volume V in cm.^{3}.

TABLE 2

λ (Å^3)	ε	λ (Å^3)	ε
5·63	0·56	12·80	0·61
6·95	0·58	13·40	0·61
8·97	0·60	13·99	0·61
9·65	0·60	14·60	0·61
10·33	0·61	15·27	0·61
10·91	0·61	15·85	0·61
11·39	0·63	17·09	0·61
11·62	0·65	18·39	0·61
11·73	0·62	19·57	0·61
12·24	0·61	20·91	0·61

TABLE 3

V	p	V	p
0·00239	41·7	0·00545	54·7
0·00295	53·3	0·00570	54·2
0·00381	57·6	0·00595	53·7
0·00410	57·6	0·00621	52·9
0·00439	57·3	0·00650	51·6
0·00464	56·7	0·00674	51·1
0·00484	52·0	0·00727	49·3
0·00494	49·6	0·00788	47·3
0·00499	54·3	0·00833	45·8
0·00521	54·9	0·00890	43·9

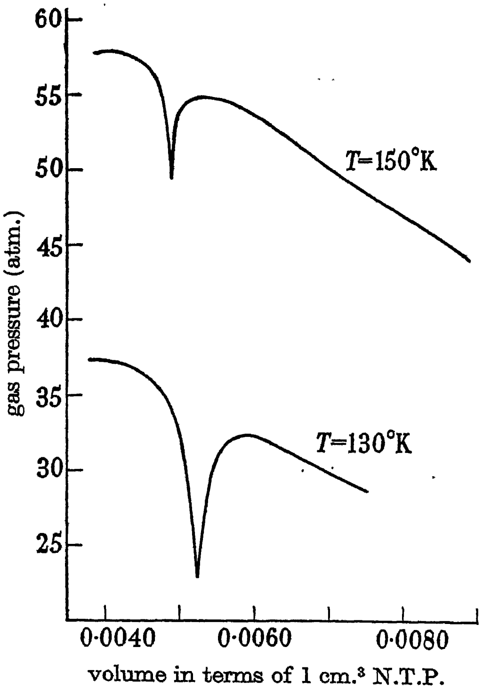


FIGURE 8

In figure 8 the result just obtained has been plotted. Similar calculations have been performed for the same gas for a temperature $T = 130°$ K and the result is shown on the same figure.

8. DISCUSSION

The general shape of the curves obtained is rather similar to the classical curve of van der Waals except that there is a sharp point instead of a smooth minimum. This point corresponds to the singularity separating the gas from the liquid; it lies in the region of completely unstable states. This confirms the statement made in the introduction, that Mayer's interpretation, namely that the singularity represents directly the point of condensation, or in other words the equilibrium of liquid and gas is not right. The latter has in fact to be found by the usual thermodynamical construction (Maxwell).

If the curves are examined in more detail it is obvious that the first one corresponds indeed almost to the critical point as the kink is very small. In the second curve the range of unstable states is considerably larger. A conspicuous feature is the fact that the curves for small volumes do not ascend steeply but have a maximum and fall again. This may be due to the crudeness of the approximation (i.e. the fact of having taken into account only the first three roots of the transcendental equation). But it seems to be quite possible that it is an indication of another range of unstable states separating the liquid from the solid. This has to be investigated.

Concerning the absolute values, one can compare the experimental critical point with the situation of the maximum in the gas region, calculated for the experimental critical temperature, figure 8. The result is given now:

	experimental	theoretical
p	50 atm.	54 atm.
V	0·0034 cm.^{3}	0·0052 cm.^{3}

The agreement is very good for the pressure and sufficiently good for the volume, and it must be kept in mind that in all previous theories, for instance that of Lennard-Jones, the deviation was still greater.

The calculations in this paper are only intended to show that one can obtain an equation of state representing the liquid and the gas states in one expression, the constants being expressed in terms of the atomic forces.

Exact numerical prediction would need a very great amount of computation. It seems more important to use the same method for calculating the radial distribution function which can directly be compared with X-ray experiments. This will be done in a following paper.

The author would here gratefully acknowledge the debt he owes to Professor Max Born, who suggested the subject of this work and gave the author invaluable advice and encouragement. He is also indebted to Dr H. S. Green for many helpful suggestions.

REFERENCES

Born, M. & Green, H. S. 1946 *Proc. Roy. Soc.* A, **188**, 10.
Born, M. & Green, H. S. 1947*a* *Proc. Roy. Soc.* A, **190**, 455.
Born, M. & Green, H. S. 1947*b* *Proc. Roy. Soc.* A, **191**, 168.
Born, M. & Fuchs, K. 1938 *Proc. Roy. Soc.* A, **166**, 391.
De Boer, J. 1940 *Contribution to the theory of compressed gases.* Amsterdam Dissertation.
Green, H. S. 1947 *Proc. Roy. Soc.* A, **189**, 103.
Kahn, B. 1938 Utrecht Dissertation.
Kirkwood, J. G. & Monroe, E. 1941 *J. Chem. Phys.* **9**, 514.
Kirkwood, J. G. & Boggs, E. M. 1942 *J. Chem. Phys.* **10**, 394.
Lennard-Jones, J. E.. 1924 *Proc. Roy. Soc.* A, **106**, 463.
Mayer, J. E. 1937 *J. Chem. Phys.* **5**, 67, 74; **6**, 87, 101.
Montroll, E. W. & Mayer, J. E. 1941 *J. Chem. Phys.* **9**, 626.
Ursell, H. D. 1927 *Proc. Camb. Phil. Soc.* **23**, 685.
Whittaker, E. T. & Robinson, G. 1937 *The calculus of observations.* London: Blackie and Son.

Some theoretical considerations on the dynamic properties of plastics

BY N. F. ASTBURY

(*Communicated by W. T. Astbury, F.R.S.—Received* 29 *June* 1948)

There is a marked similarity between the mechanical behaviour of plastics and the electric behaviour of certain dielectrics. This provides a starting-point for the derivation of a mechanical stress-strain relationship for plastic materials, from which expressions are deduced for the velocity of sound in, and for the transmission and reflexion of sound by, plastic plates. The effect of absorption on departures from Rayleigh's theory is indicated, and limited experimental data are reported.

1. INTRODUCTION

The transmission of elastic waves in an infinite slab of homogeneous material which obeys Hooke's law is completely determined in terms of two elastic constants and the density of the material (Love 1944). It is convenient to combine the two elastic constants to give one or other of two elastic moduli—the 'dilatational modulus', which governs the transmission of compressional waves, and the 'shear modulus', which governs the transmission of rotational or shear waves. On the basis of Hooke's law, which, of course, postulates a direct linear relationship between stress and strain, these moduli remain constant throughout the whole frequency range, so that the material shows neither dispersion nor absorption.

The reflexion of elastic waves from, and the transmission of elastic waves through, large parallel-sided slabs of such material have been discussed in some detail by Reissner (1938), who shows that, in general, the reflexion and transmission characteristics are determined by both moduli and the density. The classical analysis of Rayleigh (1929), as, indeed, is pointed out by the author, applies only to *fluids*;

his results are applicable to solids only when the conditions are such as to preclude shear stresses in the material, which, in general, can be taken to be the case for normal incidence.

Experimental work during the war on mechanical and acoustic properties of certain plastic materials showed, as is indeed now well known, that these substances do not obey Hooke's law; in particular, 'Perspex' (a methyl methacrylate polymer) shows considerable departures from the simple linear stress-strain relationship. Further, the classical relationship between rigidity modulus and Young's modulus does not hold. The present paper places on record some theoretical considerations on the dynamic properties of plastics and treats the problem of the transmission and reflexion of sound normally incident on plates of non-Hookean properties.

2. PRELIMINARY CONSIDERATIONS

There is a very close analogy between the mechanical behaviour of plastics and the electrical behaviour of certain dielectrics, and the preliminary arguments following will be familiar to those acquainted with the theories of dielectric hysteresis and absorption.

When an electric stress is applied to a dielectric there is an initial sudden rush of current (the normal 'charging current') which is followed by the so-called 'anomalous' current, decaying with time, which may persist for hours or even days, and which does appear ultimately to reach a limit. This anomalous current can be drawn from the dielectric during the process of discharge, and from its integrated value an apparently already high value of dielectric constant (or low electric elasticity) can be deduced. In exactly the same way, when a mechanical stress is applied to a plastic material, there is an initial sudden displacement (it is preferred at this stage to use the rather loose description 'sudden' rather than the more specific 'instantaneous') which is followed by a sort of flow which appears to go on for hours, decaying gradually so that the ultimate strain of the material does appear to reach a finite limit. The effect is very easy to demonstrate: in figure 1 is shown a curve recording the twist of a Perspex rod after the sudden application of a torque.

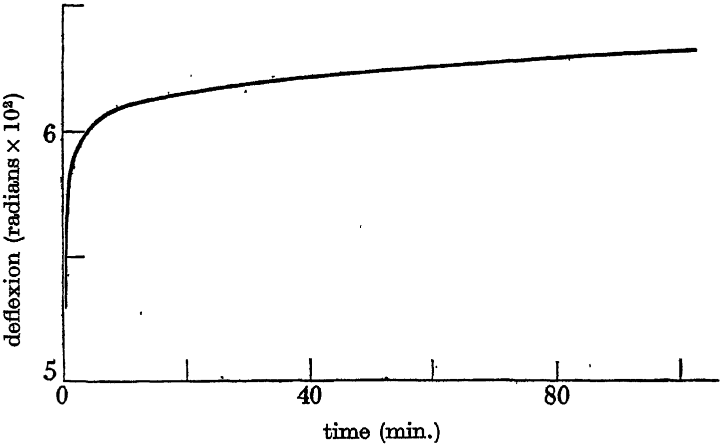


FIGURE 1. Strain curve for Perspex under constant stress.

It is well known that the application of an alternating electric stress to a dielectric exhibiting the 'anomalous' behaviour outlined above leads to hysteresis and the absorption of power. Simple mechanical experiments on the cyclic stressing of a Perspex bar or rod show at once that the corresponding phenomenon is present in this case also; figure 2 shows a hysteresis loop obtained for a slow alternating stress applied to a Perspex rod.

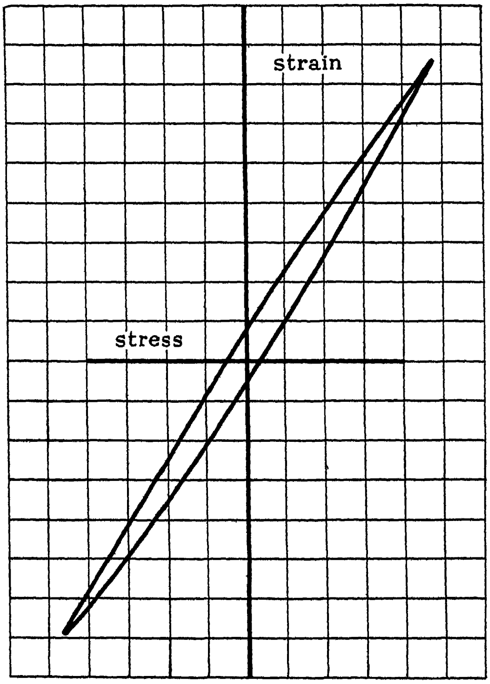


FIGURE 2. Hysteresis loop for Perspex rod.

It is reasonable to suppose, therefore, that there would be a close formal similarity between any theory put forward to explain the mechanical behaviour of plastics and any theory of dielectric absorption. We may start in the most general way possible by saying that the plastic flow or displacement, ξ_p, is simply a function of time, and we write

$$\xi_p \propto f(t), \tag{2·1}$$

where the factor of proportionality involves the stress. We further suppose that at $t = 0$, $f(t) = 0$, while for $t = \infty$ we suppose that $f(t)$ is finite. In order to discover how ξ_p changes with a changing stress, we have to make the fundamental assumption that a superposition principle is operative. This means that the increments of stress are all additive. If by F_u we denote the stress at time u, then the value of the increment of plastic flow at this instant may be written

$$\delta\xi_p = Af(t-u)\frac{dF_u}{du}\delta u, \tag{2·2}$$

and the total plastic flow in time t

$$\xi_p = A\int_{-\infty}^{t} \frac{dF_u}{du} f(t-u)\, du, \tag{2·3}$$

where A is a constant.

If the stress is alternating and of the form

$$F_u = F_0 \sin \omega u,$$

we have

$$\xi_p = A \int_{-\infty}^{t} \omega F_0 \cos \omega u \, f(t-u)\, du,$$

and changing the variable to $z = t-u$, we have

$$\begin{aligned} \xi_p &= A\omega F_0 \sin \omega t \int_0^\infty f(z) \sin \omega z \, dz + A\omega F_0 \cos \omega t \int_0^\infty f(z) \cos \omega z \, dz \\ &= \omega F_0 [S \sin \omega t + C \cos \omega t], \end{aligned} \tag{2·4}$$

where

$$S = A \int_0^\infty \sin \omega z f(z)\, dz, \quad C = A \int_0^\infty \cos \omega z f(z)\, dz.$$

So long, then, as the integrals C and S do not vanish, there is always a phase lag between the plastic flow and the stress. It follows, therefore, that there is a definite loss of power in any cyclic operation, which is easily shown to be proportional to the integral C; the progression of periodic waves through the material must therefore be accompanied by attenuation.

It is thus easy to see in general terms how absorption and hysteresis arise when we depart from Hooke's law; the next stage in the development of the theory must be the establishment of a more general stress-strain relationship.

3. General stress-strain relationship

It is no longer possible, for plastic substances, to think in terms of a simple elastic modulus, defined as the ratio of stress to strain. In electrotechnics, it is easy to determine experimentally the values of S and C separately, and we speak of effective capacitance (or dielectric constant) in terms of S and of power factor in terms of C. It is not obvious that we can do this in the mechanical case; we must either measure r.m.s. values and define our modulus in terms of $\sqrt{(S^2 + C^2)}$, or regard the relationship between stress and strain as expressed in the form of a complex operator, or more generally by a differential or integral equation. It is the last method which we shall try to adopt in the present paper.

We have first to decide whether the strain of a plastic material is wholly a function of time or whether it is partly elastic, and in the absence of reliable experimental evidence to guide us we can only make certain assumptions. We know, for instance, that in the case of materials like pitch, there appears to be no purely elastic property; for such materials Maxwell long ago proposed the relationship

$$\mu \frac{d\xi}{dt} = \frac{dF}{dt} + \alpha F, \tag{3·1}$$

where μ and α are constants. We cannot accept this form, for clearly, if F is constant, ξ increases linearly without limit, and this we do not believe to be the case with plastics. We shall, therefore, assume that the deformation of substances such as Perspex is partly purely elastic and partly plastic. The sudden application of a stress to such a material must therefore result in an *instantaneous* strain, followed by a plastic strain dying away with time. While simple mechanical experiments cannot be said to have proved this to be the case absolutely, at any rate the assumption is not inconsistent with the observed facts.

We write, then, the total strain at any time t as

$$\xi_t = \xi_i + \xi_p f(t), \tag{3·2}$$

where ξ_i is the instantaneous strain and ξ_p is the limiting value of the plastic strain. Each is assumed to bear a simple relationship to stress. The function $f(t)$ must be zero for $t = 0$ and unity for $t = \infty$. Under a stress which is assumed to be a continuous function of time, the superposition principle is supposed to operate, and we write, as in § 2,

$$\delta\xi_u = \nu_p f(t-u)\frac{dF_u}{du}du,$$

where ν_p is the constant expressing the relation between ξ_p and stress.

For the total strain we find

$$\begin{aligned}\xi_t &= \xi_i + \nu_p \int_{-\infty}^{t} f(t-u)\frac{dF_u}{du}du \\ &= \xi_i + \nu_p\Big[f(t-u)F_u\Big]_{-\infty}^{t} - \nu_p\int_{-\infty}^{t}\frac{d}{du}f(t-u)F_u du,\end{aligned}$$

or, since $f(t-u) = 0$ for $u = t$ and $F_u = 0$ for $t = -\infty$,

$$\xi_t = \nu_i F_t - \nu_p \int_{-\infty}^{t}\frac{d}{du}f(t-u)F_u du, \tag{3·3}$$

where ν_i is the constant expressing the relation between ξ_i and stress.

This is one form of the general stress-strain relationship, and it is essentially similar to that proposed by Boltzmann (1876) and recently quoted by Zener (1946); it corresponds also to that proposed by Pellat (1926) for the behaviour of dielectrics. We have written ξ_i as $\nu_i F_t$; the elastic behaviour of the material can be regarded as being governed not by one modulus only but by two constants ν_i and ν_p and a function of time as yet unspecified.

A simple form for f is

$$f(t) = 1 - e^{-\alpha t}. \tag{3·4}$$

This form expresses the condition that the material shall exhibit a form of relaxation characterized by a time constant $1/\alpha$. The simple mechanical experiments already described show that this form is applicable over only a limited range of time, which suggests that the behaviour is characterized not by one time constant but by a whole range of time constants. This is known also to be the case in dielectric theory, and at least two other forms of f have been proposed to take this into account. Schweildler (1926) has suggested that a function involving a summation $\Sigma\nu_p e^{-\alpha t}$ should be used, where every α is associated with a definite ν_p; Wagner (1926) has proposed a Gaussian distribution, in which the summation is replaced by an integral of the form

$$\int_{-\infty}^{+\infty}\exp(-b^2z^2 - \alpha_0 t e^{-z})dz,$$

where $z = \log \alpha_0/\alpha$.

For the purposes of the present paper we shall consider only the simple form (3·4), and in justification of this the following arguments are put forward. The behaviour of the material under alternating stresses can only be significantly affected by those relaxation modes, the time constants of which are comparable with the periods of the alternating stresses. The law proposed holds over a range of time comparable to two or three octaves in the frequency range. It is therefore suggested that analysis based on the simple form given in (3·4) will give an approximate picture of the behaviour of the substance over a restricted frequency range.

With this law the general stress-strain relationship (3·3) becomes

$$\xi_t = \nu_i F_t + \nu_p \alpha \int_{-\infty}^{t} \mathrm{e}^{-\alpha(t-u)} F_u \, du,$$

or

$$\xi_t \mathrm{e}^{\alpha t} = \nu_i F_t \mathrm{e}^{\alpha t} + \nu_p \alpha \int_{-\infty}^{t} \mathrm{e}^{\alpha u} F_u \, du.$$

Thus ξ_t and F_t must satisfy the differential equation

$$\frac{d\xi_t}{dt} + \alpha \xi_t = \nu_i \frac{dF_t}{dt} + \alpha(\nu_i + \nu_p) F_t. \tag{3·5}$$

We note that we can define two limiting values of elastic modulus corresponding to infinite and zero frequency. At infinite frequency the modulus reaches a non-complex limit equal to $1/\nu_i$; this will be called μ_∞. For zero frequency the modulus again has a limiting non-complex value equal to $1/(\nu_i + \nu_p)$; this will be called μ_0. It must approximate closely to what we may call the measurable static modulus of the material. The general stress-strain relationship, over a restricted frequency range, then becomes

$$\mu_\infty \left(\frac{d\xi}{dt} + \alpha \xi \right) = \frac{dF}{dt} + \alpha \frac{\mu_\infty}{\mu_0} F_t. \tag{3·6}$$

Using this relationship in place of Hooke's law, it is proposed to deduce expressions for the velocity of sound, and to examine the problem of reflexion and transmission by thin plates.

4. VELOCITY OF SOUND

We have already shown that departures from Hooke's law lead to absorption; the displacement, y, in a progressive wave through the material can therefore be represented by

$$y = y_0 \exp\left[j\omega \left\{ t - \frac{x}{c}(1 - jb) \right\} \right], \tag{4·1}$$

where c = velocity and b = attenuation factor. The strain, ξ, is given by

$$\xi = \frac{\partial y}{\partial x} = \frac{-j\omega(1-jb)}{c} y,$$

and also

$$\frac{\partial \xi}{\partial t} = j\omega \xi.$$

Thus

$$\frac{\partial F}{\partial t} + \alpha \frac{\mu_\infty}{\mu_0} F = \mu_\infty \xi (\alpha + j\omega),$$

whence
$$F=\frac{-j\omega\mu_\infty(1-jb)\,(\alpha+j\omega)}{c\left\{\alpha\frac{\mu_\infty}{\mu_0}+j\omega\right\}}\,y.$$

We have also, if ρ be the density of the material,

$$\frac{\partial F}{\partial x}=\rho\frac{\partial^2 y}{\partial t^2},$$

so that, if $c_\infty^2=\mu_\infty/\rho$,
$$(1-jb)^2\,(\alpha+j\omega)=\frac{c^2}{c_\infty^2}\left\{\alpha\frac{\mu_\infty}{\mu_0}+j\omega\right\}. \tag{4·2}$$

Solving for c and b, we get

$$c=\sqrt{\left(\frac{2(\alpha^2+\omega^2)}{\alpha^2\frac{\mu_\infty}{\mu_0}+\omega^2+\sqrt{\left\{\left(\alpha^2\frac{\mu_\infty}{\mu_0}+\omega^2\right)^2+\left(\frac{\mu_\infty}{\mu_0}-1\right)^2\alpha^2\omega^2\right\}}}\right)}\sqrt{\left(\frac{\mu_\infty}{\rho}\right)} \tag{4·3}$$

$$=\sqrt{\left(\frac{2(\alpha^2+\omega^2)}{\alpha^2+\frac{\mu_0}{\mu_\infty}\omega^2+\sqrt{\left\{\left(\alpha^2+\frac{\mu_0}{\mu_\infty}\omega^2\right)^2+\left(1-\frac{\mu_0}{\mu_\infty}\right)^2\alpha^2\omega^2\right\}}}\right)}\sqrt{\left(\frac{\mu_0}{\rho}\right)}, \tag{4·4}$$

$$b=\frac{\mu_\infty-\mu_0}{\mu_\infty}\,\frac{\alpha\omega}{\alpha^2+\frac{\mu_0}{\mu_\infty}\omega^2+\sqrt{\left\{\left(\alpha^2+\frac{\mu_0}{\mu_\infty}\omega^2\right)^2+\left(1-\frac{\mu_0}{\mu_\infty}\right)^2\alpha^2\omega^2\right\}}}. \tag{4·5}$$

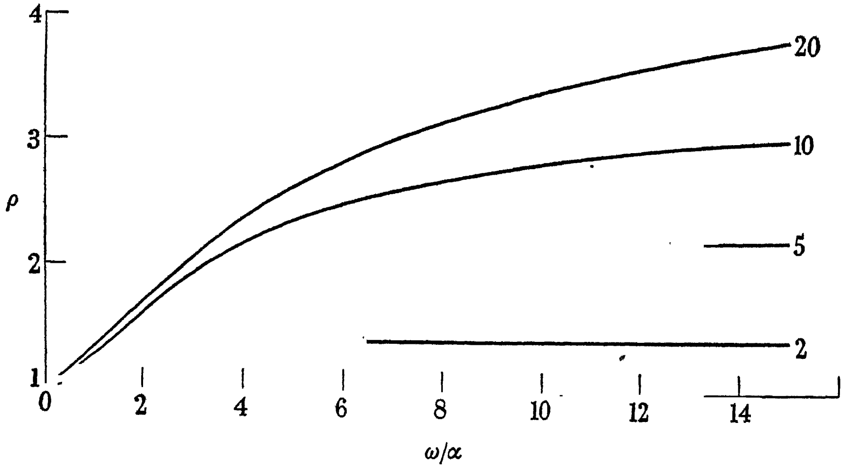


Figure 3. Variation of velocity of sound with frequency. The numbers against the curves are the values of μ_∞/μ_0. ρ = (velocity at angular frequency ω)/(velocity at zero frequency).

In equations (4·3) and (4·4) the quantities $\sqrt{(\mu_\infty/\rho)}$ and $\sqrt{(\mu_0/\rho)}$ represent the limiting values of the velocity of sound. The latter is the velocity, c_0, as calculated from static measurements; the function of ω which multiplies it in (4·3) varies from unity at $\omega = 0$ to $\surd(\mu_\infty/\mu_0)$ at $\omega = \infty$. The attenuation factor b† is zero at $\omega = 0$ and $\omega = \infty$ and reaches a maximum value of $(c_\infty - c_0)/(c_\infty + c_0)$ for $\omega/\alpha = \surd(\mu_\infty/\mu_0)$. The actual curves of c/c_0 and b are sketched out in figures 3 and 4.

It is clear that a critical test of the theory would be provided by absorption measurements.

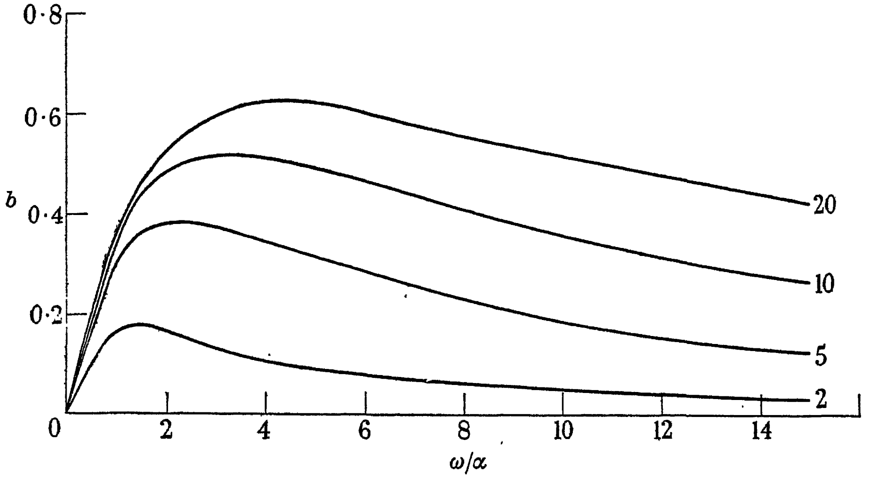


FIGURE 4. Variation of attenuation factor with frequency. Loss in thickness $t = 1 - e^{(-2\pi b t)/\lambda}$. The numbers against the curves are the values of μ_∞/μ_0.

5. REFLEXION AND TRANSMISSION OF SOUND INCIDENT NORMALLY ON A PLANE PARALLEL-SIDED LAMINA OF PLASTIC MATERIAL

The comparatively simple case of a plane wave incident normally on an infinite plane parallel-sided lamina can be dealt with on the present theory.

Let c_1 be the velocity of the wave in the surrounding medium, which we take to be normal in its properties. Let c_2 be the velocity of sound in the plastic lamina, of which the thickness is l, and let b be the attenuation factor. We define also

ρ_1 = density of medium,

ρ_2 = density of plastic,

μ_1 = modulus of medium, so that

$c_1 = \surd(\mu_1/\rho_1)$,

μ_2 = upper limit of modulus of lamina.

As before F and ξ, with appropriate suffixes, will be used for stress and strain respectively. The incident wave will be assumed to travel in the direction of x increasing, and the plane of the lamina will be parallel to the yz plane. To simplify the writing down of the analysis we may, since normal incidence is being considered, consider displacements only in the xz plane; the lamina will be assumed bounded by the planes $x = 0$ and $x = l$. The total displacements in the medium on the incident side, in the lamina itself, and in the medium on the emergent side, may then be written as

---

† The absorption coefficient, which is the quantity multiplied on x to give the real part of the exponent in (4·1), is of course $2\pi b/\lambda$, where λ = wave-length.

$$Z_1 = z_1 \exp\left\{j\omega\left(t - \frac{x}{c_1}\right)\right\} + z_2 \exp\left\{j\omega\left(t + \frac{x}{c_1}\right)\right\}, \tag{5·1}$$

$$Z_2 = z_3 \exp\left[j\omega\left\{t - \frac{x}{c_2}(1 - jb)\right\}\right] + z_4 \exp\left[j\omega\left\{t + \frac{x}{c_2}(1 - jb)\right\}\right], \tag{5·2}$$

$$Z_3 = z_5 \exp\left\{j\omega\left(t - \frac{x}{c_1}\right)\right\}. \tag{5·3}$$

The incident wave is thus $$z_1^* \equiv z_1 \exp\left\{j\omega\left(t - \frac{x}{c_1}\right)\right\}, \tag{5·4}$$

and the reflected wave is $$z_2^* \equiv z_2 \exp\left\{j\omega\left(t + \frac{x}{c_1}\right)\right\}, \tag{5·5}$$

and the emergent wave is $$z_5^* \equiv z_5 \exp\left\{j\omega\left(t - \frac{x}{c_1}\right)\right\}. \tag{5·6}$$

The waves in the plate are

$$z_3^* \equiv z_3 \exp\left[j\omega\left\{t - \frac{x}{c_2}(1 - jb)\right\}\right], \tag{5·7}$$

$$z_4^* \equiv z_4 \exp\left[j\omega\left\{t + \frac{x}{c_2}(1 - jb)\right\}\right]. \tag{5·8}$$

At the boundaries of the lamina we assume that the displacement is the same in both the medium and the lamina, and we also equate the normal components of stress at these points. The quantities z are all complex.

In order to find the stress in the lamina at the points $x = 0$ and $x = l$ we make use of equation (3·6). The strain, ξ, in the lamina is given by

$$\xi = \frac{\partial Z_2}{\partial x} = -\frac{j\omega}{c_2}(1 - jb)(z_3^* - z_4^*),$$

and $$\frac{\partial \xi}{\partial t} = \frac{\omega^2}{c_2}(1 - jb)(z_3^* - z_4^*).$$

Thus, remembering that μ_2 is identical with the μ_∞ of (3·6),

$$\frac{\partial F}{\partial t} + \alpha\frac{\mu_2}{\mu_0}F = -\mu_2\frac{j\omega}{c_2}(1 - jb)(\alpha + j\omega)(z_3^* - z_4^*),$$

giving $$F = \frac{-j\omega\mu_2(1 - jb)(\alpha + j\omega)}{c_2\left(j\omega + \dfrac{\mu_2}{\mu_0}\alpha\right)}(z_3^* - z_4^*).$$

Thus the values of F at $x = 0$ and $x = l$ are

$$F_0 = \frac{-j\omega\mu_2(1 - jb)(\alpha + j\omega)}{c_2\left(j\omega + \dfrac{\mu_2}{\mu_0}\alpha\right)}(z_3 - z_4)\,e^{j\omega t},$$

$$F_l = \frac{-j\omega\mu_2(1 - jb)(\alpha + j\omega)}{c_2\left(j\omega + \dfrac{\mu_2}{\mu_0}\alpha\right)}\left(z_3 \exp\left\{-j\omega\frac{l}{c_2}(1 - jb)\right\} - z_4 \exp\left\{j\omega\frac{l}{c_2}(1 - jb)\right\}\right)e^{j\omega t}.$$

Thus the boundary equations become, remembering that the stresses in the medium are given by

$$\mu_1\frac{\partial Z_1}{\partial x} \quad \text{and} \quad \mu_1\frac{\partial Z_3}{\partial x},$$

$$z_1+z_2 = z_3+z_4, \tag{5·9}$$

$$(p+jq)(z_1-z_2) = z_3-z_4, \tag{5·10}$$

$$\frac{z_5}{u} = \frac{z_3}{v}+vz_4, \tag{5·11}$$

$$(p+jq)\frac{z_5}{u} = \frac{z_3}{v}-vz_4, \tag{5·12}$$

where

$$\left.\begin{aligned} p+jq &= \frac{\mu_1 c_2}{\mu_2 c_1}\frac{j\omega+\dfrac{\mu_2}{\mu_0}\alpha}{(1-jb)(\alpha+j\omega)},\\ u &= \exp(j\omega l/c_1),\\ v &= \exp\{j\omega l(1-jb)/c_2\}. \end{aligned}\right\} \tag{5·13}$$

For the reflected wave we find

$$\frac{z_2}{z_1} = \frac{1-(p+jq)^2}{1+(p+jq)^2-2(p+jq)\dfrac{1+v^2}{1-v^2}}, \tag{5·14}$$

and for the transmitted wave

$$\frac{z_5}{z_1} = \frac{4uv(p+jq)}{(1-v^2)\left\{1+(p+jq)^2-2(p+jq)\dfrac{1+v^2}{1-v^2}\right\}}. \tag{5·15}$$

From equation (4·2) we have

$$(1-jb)^2(\alpha+j\omega) = \frac{c_2^2}{c_\infty^2}\left\{\alpha\frac{\mu_2}{\mu_0}+j\omega\right\},$$

so that

$$\begin{aligned} p+jq &= \frac{\mu_1 c_2}{\mu_2 c_1}\frac{j\omega+\dfrac{\mu_2}{\mu_0}\alpha}{(1-jb)(\alpha+j\omega)}\\ &= \frac{c_1^2\rho_1 c_2}{c_\infty^2 c_1\rho_2}(1-jb)\frac{c_\infty^2}{c_2^2}\\ &= \frac{c_1\rho_1}{c_2\rho_2}(1-jb)\\ &\equiv s(1-jb), \end{aligned} \tag{5·16}$$

where s is the ratio of the velocity-density products *at the frequency of observation*. Thus $p = s$, $q = -bs$. Further

$$\begin{aligned} \frac{1+v^2}{1-v^2} &= \frac{\exp\{j\omega l(1-jb)/c_2\}+\exp\{-j\omega l(1-jb)/c_2\}}{\exp\{j\omega l(1-jb)/c_2\}-\exp\{-j\omega l(1-jb)/c_2\}}\\ &= j\cot\omega l(1-jb)/c_2\\ &= j\frac{\sin\phi+j\sinh b\phi}{\cosh b\phi-\cos\phi}, \end{aligned}$$

where $\phi = 2\omega l/c_2$. Thus for the reflected wave we get

$$\frac{z_2}{z_1} = \frac{1 - s^2(1-jb)^2}{1 + s^2(1-jb)^2 - 2s(1-jb)\,j\,\dfrac{\sin\phi + j\sinh b\phi}{\cosh b\phi - \cos\phi}}$$

$$= \frac{1 - s^2(1-b^2) + 2jbs^2}{1 + s^2(1-b^2) - 2sf_1(\phi) - 2js\{bs + f_2(\phi)\}},$$

where

$$f_1(\phi) = \frac{b\sin\phi - \sinh b\phi}{\cosh b\phi - \cos\phi}, \tag{5·17}$$

$$f_2(\phi) = \frac{\sin\phi + b\sinh b\phi}{\cosh b\phi - \cos\phi}. \tag{5·18}$$

Taking r.m.s. values we find for the reflexion coefficient, R,

$$R = \sqrt{\left(\frac{\{1 - s^2(1-b^2)\}^2 + 4b^2s^4}{[1 + s^2(1-b^2) - 2sf_1(\phi)]^2 + 4s^2[bs + f_2(\phi)]^2}\right)}. \tag{5·19}$$

When $b = 0$, we find $f_1(\phi) = 0$ and $f_2(\phi) = \cot\frac{1}{2}\phi$, so that R then becomes

$$R_0 = \sqrt{\left(\frac{(1-s^2)^2}{[1+s^2]^2 + 4s^2\cot^2\frac{1}{2}\phi}\right)}$$

$$= \frac{1/s - s}{\sqrt{\{(1/s+s)^2 + 4\cot^2\frac{1}{2}\phi\}}},$$

which is identical with Rayleigh's classical result. By similar reduction it can be shown that the transmission coefficient, T (which is the r.m.s. value of z_5/z_1), is given by

$$T = \sqrt{\left(\frac{8s^2(1+b^2)}{(\cosh b\phi - \cos\phi)\{[1 + s^2(1-b^2) - 2sf_1(\phi)]^2 + 4s^2[bs + f_2(\phi)]^2\}}\right)}, \tag{5·20}$$

and when $b = 0$ this reduces to

$$T_0 = \sqrt{\left(\frac{8s^2}{2\sin^2\frac{1}{2}\phi\{(1+s^2)^2 + 4s^2\cot^2\frac{1}{2}\phi\}}\right)}$$

$$= \frac{2}{\sqrt{\{4\cos^2\frac{1}{2}\phi + (1/s+s)^2\sin^2\frac{1}{2}\phi\}}},$$

which is identical with Rayleigh's classical result.

6. THE DEGENERATION OF THE CLASSICAL RESULT DUE TO ABSORPTION

It is clear from the nature of the expressions derived above that no significant change of property will occur unless ω/α is of the order of unity. Physically, this means that the material must exhibit relaxation frequencies comparable with the frequency of observation, and therefore must display absorption within the working range. The foregoing results have been considered numerically for values of α lying between 10 and 40 kc./sec.

In figure 5 are plotted reflexion coefficients, computed from equation (5·19), for Perspex. The basic data are derived from static measurements and assumed values of α (namely, 10, 20 and 40 kc./sec.) have been taken together with an assumed value of 10 for the ratio μ_∞/μ_0. The Rayleigh curve, computed from the static data, is dotted in for reference. The profound difference is at once evident.

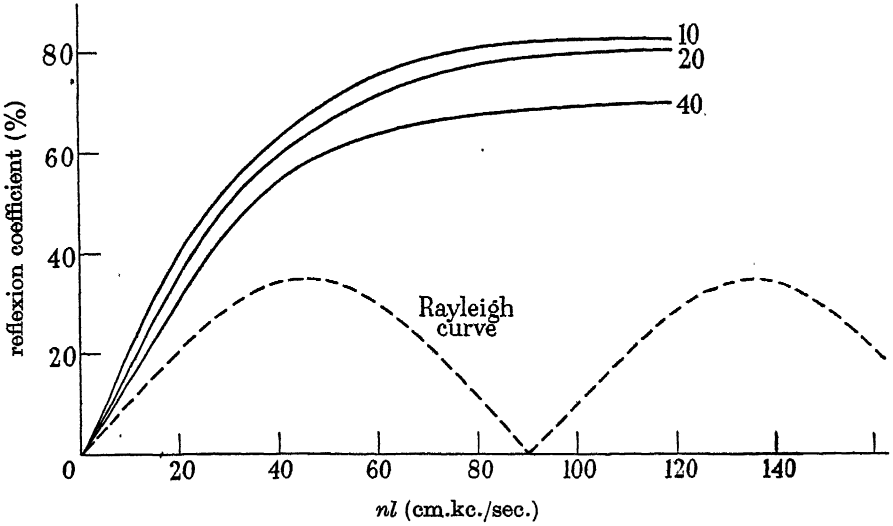


FIGURE 5. Degeneration of the Rayleigh curve for Perspex ($\mu_\infty/\mu_0 = 10$). $n = \omega/2\pi$. The numbers against the curves are the values of α.

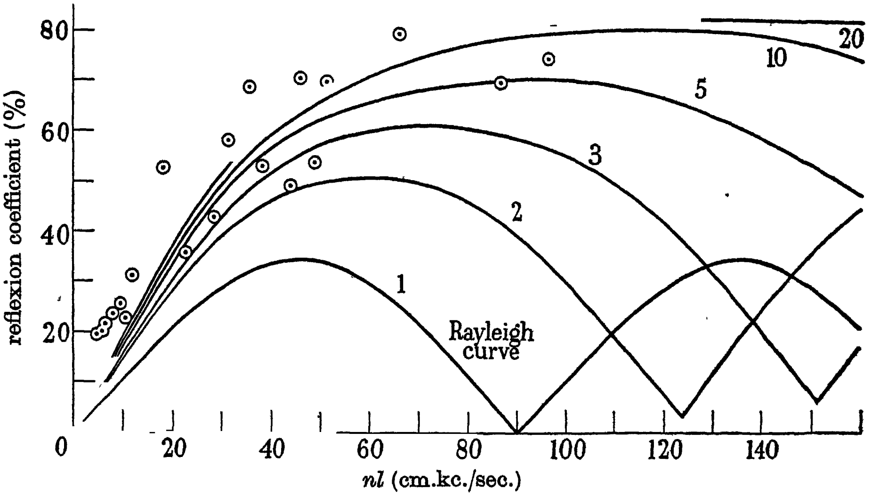


FIGURE 6. Degeneration of the Rayleigh curve for Perspex ($\alpha = 20$ kc./sec.). $n = \omega/2\pi$. The numbers against the curves are the values of μ_∞/μ_0.

In figure 6 are plotted reflexion coefficients for an assumed value of α of 20 kc./sec., with values of μ_∞/μ_0 ranging from unity (the Rayleigh case) to 20. This figure shows how the Rayleigh case can degenerate until curves like those in figure 5 are reached. It is noteworthy that quite plausible imitations of the Rayleigh curve are obtained for low values of μ_∞/μ_0; one might well be justified in regarding such curves, if found experimentally, as true Rayleigh curves within the limits of experimental error, and yet these curves differ considerably from the Rayleigh curve computed from static data.

To illustrate this point more fully, consider the first three curves of figure 6. From the first, with its zero at $nl = 90$ cm.kc./sec., we deduce $c = 1{\cdot}8 \times 10^5$ cm./sec. From the second and third, we might reasonably take c to be $2{\cdot}5 \times 10^5$ and $3{\cdot}0 \times 10^5$ cm./sec. respectively. The absorption required to produce this change can be estimated from the curves of figure 4, and obviously corresponds to value of b of the order of 0·1. This means that the absorption in a plate 1 cm. thick would be of the order of $1 - \exp(2\pi b/\lambda) = 1 - \exp(0{\cdot}06) = 6\,\%$. Thus a material exhibiting quite small absorption may have a reflexion coefficient much greater than that computed from static data.

On figure 6 are marked a number of experimentally observed points obtained with Perspex disks, of 0·125, 0·5 and 1 in. thickness, at frequencies between 10 and 40 kc./sec. The Rayleigh curve is computed from static data for this material and obviously does not nearly approach the observed points, the general disposition of which much more closely follows the higher order curves. The measurements of reflexion coefficient were made under water with a pulsed transmitter, using as a reference standard an effective total reflector in the form of an air-filled capsule of the same size and shape as the test disk, a method successfully used by Boyle & Rawlinson (1938) and other workers, and which yields for elastic materials results in close accord with the Rayleigh theory.

7. CONCLUSION

It is shown in this paper that materials such as Perspex and other plastics in their departure from Hooke's law display complex elastic moduli and may exhibit acoustic properties quite different from those predicted by the classical elastic theory. In particular, the reflexion and absorption of high-frequency sound by these materials may be much greater than is suggested by their 'static' or low-frequency elastic constants, and it is evident that the behaviour of Perspex is much more in accord with the present theory than with the classical theory. It is hoped that it will be possible for other workers to establish the existence or otherwise of absorption bands in the supersonic range and to determine the dispersion characteristics by direct velocity measurements.

This paper, which formed part of work done during the war, and was originally issued as a confidential report in 1942, is published by permission of the Chief of the Royal Naval Scientific Service. To his former colleague, Dr E. A. Alexander, the author records his warmest thanks for many stimulating discussions.

REFERENCES

Boltzmann, L. 1876 *Ann. Phys., Lpz.*, 7, 624.
Boyle, R. W. & Rawlinson, W. F. 1938 *Trans. Roy. Soc. Canad.* **22**, 55.
Love, A. E. H. 1944 *Elasticity*, 4th ed. p. 294. New York: Dover Publications.
Pellat, H. 1926 *E.R.A. Tech. Rep.* L/T. 14, p. 31.
Rayleigh, Lord 1929 *Theory of sound*, 2, 87. London: Macmillan.
Reissner, H. 1938 *Helv. Phys. Acta*, **11**, 140.
Schweidler, E. 1926 *E.R.A. Tech. Rep.* L/T. 14, p. 32.
Wagner, K. W. 1926 *E.R.A. Tech. Rep.* L/T. 14, p. 32.
Zener, C. 1946 *Metals Technology*. TP 1992. New York: *Amer. Inst. Min. and Met. Engng.*

The influence of diffusion on flame propagation

BY A. G. GAYDON AND H. G. WOLFHARD

Chemical Engineering Department, Imperial College, London

(*Communicated by Sir Alfred Egerton, Sec.R.S.—Received* 23 *July* 1948)

In stationary pre-mixed flames of hydrocarbons at pressures between 1 atm. and a few mm. the thickness of the reaction zone varies approximately inversely with the pressure, and the flame speed is independent of pressure. Marked exothermic reaction begins around 700 to 800° C, and the time of passage through the pre-heating zone is very much less than the induction period for ignition at this temperature; a purely thermal mechanism of flame propagation is inadequate. The thickness of the reaction zone is related to the distance which free hydrogen atoms may diffuse against the gas stream. Calculations are made on the relation between rates of diffusion, flame thickness and flame speed, and are compared with experimental values. It appears that diffusion of free atoms or radicals can account for flame propagation satisfactorily; the correct relation that the flame speed is nearly independent of pressure is obtained.

INTRODUCTION

While many attempts have been made to explain the propagation of a flame through a combustible mixture in terms of a purely thermal theory, it has also been pointed out, for example by Lewis & von Elbe (1934), that diffusion of chemically active species (free radicals or free atoms) may be of importance in initiating the combustion ahead of the flame front. Recently further evidence in support of this view has been accumulating. Thus Egerton & Powling (1948) have shown that the rate of flame propagation in mixtures is in many cases not influenced by the addition of 'promoters', although these affect the ignition temperature and would therefore be expected to affect the flame speed if this were governed by purely thermal processes. In flames at very low pressure, where the reaction zone is much thicker, we have obtained evidence (Gaydon & Wolfhard 1948) from absorption spectra that the concentration of OH radicals just in front of the reaction zone is much higher than that calculated thermally. The relation between flame velocity and the rate of diffusion from active centres has also been discussed by Van Tiggelen (1946) for methane-air flames. Most suggestive of all, Tanford & Pease (1947) have shown that the flame velocity in moist carbon monoxide-oxygen flames depends on the concentration of free hydrogen atoms, although oxygen atoms are ineffective, at least in this flame. They have made some calculations on the flame velocity and rate of diffusion of hydrogen atoms against the gas stream. These are important in indicating the undoubted role of the diffusion of hydrogen atoms, but their conclusion that the flame velocity should vary inversely as the fourth root of the gas pressure is inconsistent with our observations on the effect of pressure on flame speed for various hydrocarbon flames; provided tubes of sufficiently large diameter are used, to reduce wall effects, flames of acetylene appear to have an almost constant velocity down to very low pressures (1 mm. of Hg).

We believe that hot stationary flames are maintained by diffusion of free atoms and radicals back from the hot burnt gases, rather than by the creation of new ignition centres. It is the purpose of this short paper to discuss the effect of diffusion on the rate of flame propagation and on the thickness of the reaction zone.

THE RELATION BETWEEN FLAME VELOCITY AND REACTION VELOCITY

If we make some rather rough approximations it is possible to relate the rate of chemical reaction in the reaction zone of a flame with the flame velocity. The treatment given here is a development of that of Damköhler (1940). Let us consider the flame front as being divided into a pre-heating zone of thickness δ_{pr} and a reaction zone of thickness δ_r. At the division between these zones we may locate the ignition point, at which the reaction commences. The ignition point may be defined as that point before which there is no appreciable exothermic reaction.

Now let us use the symbols T_i, T_{ign}, T_r and T_f to denote the temperatures respectively initially, at the ignition point, in the reaction zone (mean value), and finally after reaction is complete.

Let n_c be the number of combustible molecules per cm.^{3} present initially, let u be the reaction velocity, and let v be the flame velocity, which will be equal to the initial rate of gas flow for a stationary flame. We shall measure distance x through the flame front from the ignition point as origin.

Now, by definition, the number of molecules consumed by reaction is

$$n_c v = \int_0^{\delta_r} u\,dx = \bar{u}\,\delta_r, \tag{1}$$

where $\bar{u}$ is the mean reaction velocity defined by this equation.

If we plot the temperature through the reaction zone against x, then, with Damköhler, we may make the approximation that the temperature gradient at the ignition point is related to the average temperature gradient through the reaction zone by a constant F, so that

$$\left(\frac{\partial T}{\partial x}\right)_{ign} = \frac{T_f - T_{ign}}{\delta_r} F. \tag{2}$$

Plots of the temperature through the reaction zone indicate that F has a value between 1 and 2.

Now taking λ_{ign} = heat conductivity coefficient at the ignition point, $\bar{c}_p$ = mean specific heat at constant pressure in the pre-heating zone, and ρ = density in the initial state, we may express the heat current at the ignition point as

$$\lambda_{ign}\left(\frac{\partial T}{\partial x}\right)_{ign} = \bar{c}_p \rho v (T_{ign} - T_i). \tag{3}$$

Combining (2) and (3)

$$\frac{\lambda_{ign}(T_f - T_{ign})}{\delta_r} F = \bar{c}_p \rho v (T_{ign} - T_i). \tag{4}$$

Inserting the value of δ_r from (1), we obtain

$$\frac{\lambda_{ign}(T_f - T_{ign})}{n_c v/\bar{u}} F = \bar{c}_p \rho v (T_{ign} - T_i),$$

giving

$$v^2 = \frac{F \lambda_{ign} \bar{u} (T_f - T_{ign})}{\bar{c}_p \rho n_c (T_{ign} - T_i)}. \tag{5}$$

From this it should follow that the square of the flame velocity will be roughly proportional to the reaction velocity. This formula, which has sometimes been presented in rather different forms, has never been properly tested experimentally. The only directly observable quantity is the thickness of the luminous reaction zone, and we shall now proceed to examine the nature of this.

THE THICKNESS OF THE LUMINOUS REACTION ZONE

Flames of most organic fuels, especially hydrocarbons, always show intense light emission from a thin region where the combustion is taking place. This corresponds to the 'inner cone' of flames of the Bunsen-burner type. In a flame at atmospheric pressure it is very difficult accurately to measure the thickness of this luminous reaction zone because it is so thin. With flames at low pressure the dimensions are increased. It has been found experimentally (Wolfhard 1943) that over a wide range of pressure the thickness of the luminous zone varies inversely as the gas pressure. By measuring the thickness at various pressures it has been possible to extrapolate back to atmospheric pressure with fairly high accuracy. Approximate flame velocities at atmospheric pressure have also been determined in the usual way from a knowledge of the gas flow and dimensions of the cone of the flame.

The following are the values of the thickness of the luminous reaction zone, δ_l, and the flame velocity v for various mixtures at atmospheric pressure:

mixture	δ_l (cm.)	v (cm./sec.)
$C_2H_2 + 2\frac{1}{2}O_2$	0·0021*	800
$C_2H_2 + 2\frac{1}{2}O_2 + 9{\cdot}4N_2$	0·0065	150
$C_2H_2 + 2\frac{1}{2}O_2 + 9{\cdot}4A$	0·0045	240
$C_4H_{10} + 6\frac{1}{2}O_2 + 24{\cdot}5N_2$	0·02	40

* This value has been recalculated and is rather smaller than given previously.

These values are for the luminous zone, and are measured between the points of maximum rate of change of luminosity on each side of the zone, i.e. between 'contrast zones'. The real thickness of the luminous region is, of course, greater than this, but this is the only value which is readily measurable.

These measurements are for the luminous part of the reaction zone, and we do not know, *a priori*, whether this corresponds to the reaction zone of the flame, or whether it includes part or all of the pre-heating zone, or whether it corresponds to only a part of the reaction zone. In the flames at low pressure, with a relatively thick reaction zone, measurements have been reported (Klaukens & Wolfhard 1948) on the temperature distribution through the flame front as determined with a thermocouple. Figure 1 shows curves of thermocouple reading against height above the burner top for oxy-acetylene and for three acetylene-air mixtures. In these the positions, relative to the burner top, of the luminous zones are indicated, the full line indicating the measured portion between contrast zones, and the broken curve the approximate region over which some luminosity was detectable. These curves show that the top of the luminous zone corresponds fairly well with the attainment of maximum temperature, and therefore serves to indicate the top of the reaction zone quite well.

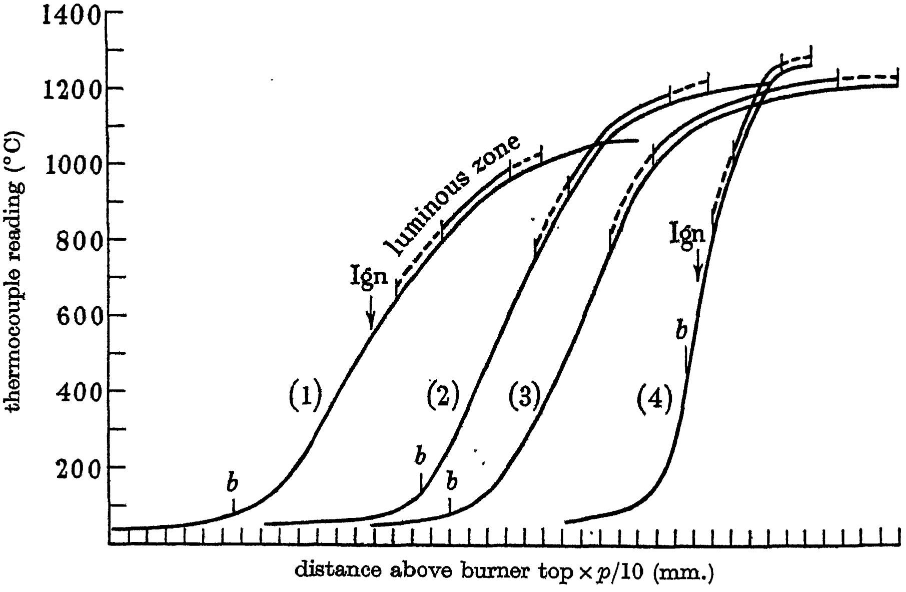


FIGURE 1. Curves of thermocouple reading (uncorrected) against distance above the burner top for four flames. The distance above the burner top is reduced to the value for a flame at 10 mm. pressure in each case to make the curves comparable, by multiplying by the pressure, p, divided by 10. The position of the burner top is indicated in each case by the letter b. The ignition points for curves (1) and (4) are indicated by 'Ign' and are taken from figure 2. The regions of luminosity are indicated. (1) 1·5 cm.^{3}/sec. C_2H_2, 16·5 cm.^{3}/sec. air. Pressure 14·8 mm. (2) 1·61 cm.^{3}/sec. C_2H_2, 16·5 cm.^{3}/sec. air. Pressure 12·0 mm. (3) 1·97 cm.^{3}/sec. C_2H_2, 16·5 cm.^{3}/sec. air. Pressure 11·4 mm. (4) 1·9 cm.^{3}/sec. C_2H_2, 12·2 cm.^{3}/sec. O_2. Pressure 4·8 mm. All measurements are for flat flames with a burner of 34 mm. diameter.

In figure 2 the beginnings of the curves (1) and (4) of figure 1 have been redrawn, but with a correction for loss of heat from the thermocouple by radiation; curves have also been calculated, and are shown by the broken lines, for the rise of temperature due to heat flow back from the burnt gases. Curves of this type have been given previously (Klaukens & Wolfhard 1948). It will be seen that the observed curve of temperature against distance through the flame front agrees closely with the calculated curve up to a temperature of around 700° C, above which the heat liberated by the exothermic combustion processes begins to become appreciable.

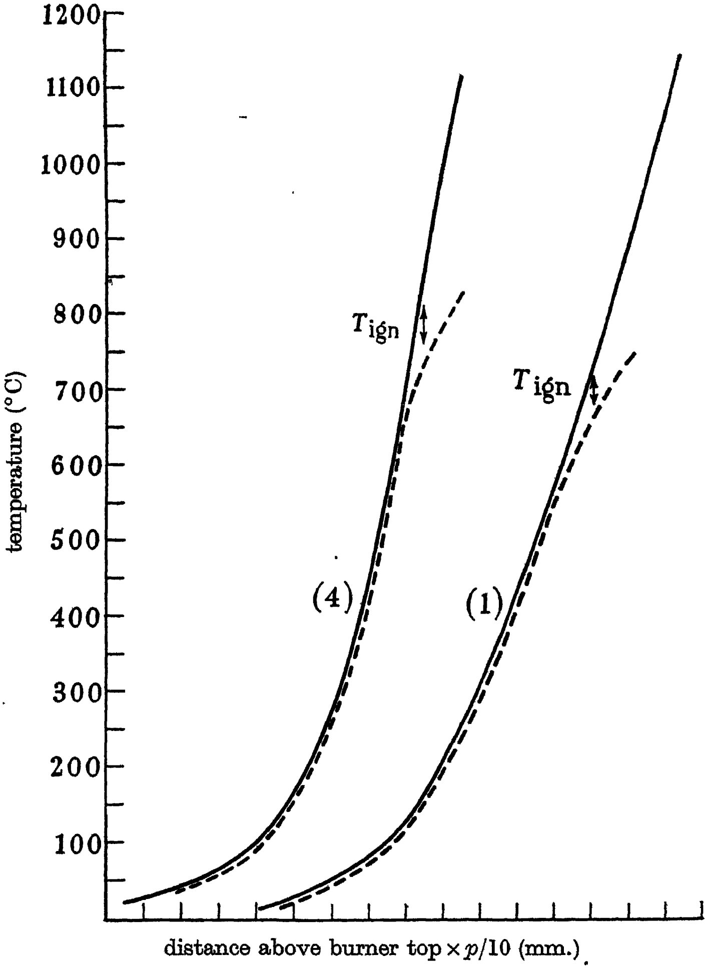


FIGURE 2. Curves (1) and (4) of figure 1 are reproduced but with the thermocouple readings corrected for radiation losses. The dotted curves are the calculated temperatures which should be attained due to thermal conduction from the hot flame gases. The ignition temperatures, T_{ign}, at which exothermic reaction becomes appreciable, are indicated by arrows.

We may thus locate the ignition point roughly as that point where the observed and calculated curves begin to show appreciable divergence. The ignition points determined in this way are indicated by vertical arrows in figure 2, and are also marked in similarly in figure 1. Again we see that the beginning of the reaction zone, as indicated by these arrows, corresponds fairly well with the points at which the first impression of light is obtained. We may draw two conclusions from this. First, the ignition point corresponds to a temperature of about 700 to 800° C. Secondly, the thickness of the reaction zone is equal to the thickness of the total luminous region, although it is somewhat larger than the region δ_l measured between the contrast zones; we suppose that the course of the reaction velocity is very similar to the light-intensity curve.

THE TIME OF PASSAGE THROUGH THE REACTION ZONE

The average time a molecule remains in the reaction zone will depend on the thickness of the zone δ_r and the average velocity through the zone. This velocity will be greater than the flame velocity, v, because of the rise of temperature. The change in the number of molecules present as a result of the chemical reaction and to a less extent as a result of dissociation will have a slight effect on the amount by which the gases expand in passing through the reaction zone. However, as an approximation we may write

$$t_r = \frac{\delta_r}{v_r} \approx \frac{\delta_r}{v} \Big/ \frac{T_r}{T_i} \approx \delta_r \Big/ \left(v \frac{T_f + T_{ign}}{2T_i} \right).$$

For a stoichiometric oxy-acetylene flame at 1 atm. $\delta_l = 0{\cdot}002$ cm., and hence δ_r will be about 0·003 cm. and t_r around 4×10^{-7} sec. The time of passage through the zone in which the temperature rises from 500 to 800° C, where we might expect the initiating reactions to occur, will be still smaller. Since we have seen that exothermic reaction is appreciable by the time a temperature of 800° is reached, and at this temperature the induction period must be very much greater than 10^{-7} sec., we see that purely thermal propagation is inadequate to account for the observations.

In the following table we give values of t_l, the time of passage through the luminous reaction zone, for some flames. These values of t_l are calculated from the observations of δ_l and will be a little less than the values of t_r. The times vary by a factor of 200 between the fastest and slowest flames studied.

mixture	t_l (sec.)
$C_2H_2 + 2\frac{1}{2}O_2$	$3{\cdot}1 \times 10^{-7}$
$C_2H_2 + 2\frac{1}{2}O_2 + 9{\cdot}4N_2$	$6{\cdot}9 \times 10^{-6}$
$C_2H_2 + 2\frac{1}{2}O_2 + 9{\cdot}4A$	$2{\cdot}6 \times 10^{-6}$
$C_4H_{10} + 6\frac{1}{2}O_2 + 24{\cdot}5N_2$	$6{\cdot}8 \times 10^{-5}$

THE RELATION BETWEEN DIFFUSION AND THE THICKNESS OF THE REACTION ZONE

The distance x travelled by a molecule by diffusion in time t is given by the usual relation*

$$x^2 = 2Dt,$$

where D is the diffusion coefficient. For flame gases D is rather difficult to evaluate because of its dependence on temperature. If we consider the diffusion of hydrogen atoms, which diffuse most rapidly, and which appear, from the work of Tanford & Pease, to be of major importance, we can make some approximate calculations. We assume an average temperature between the ignition temperature and the final temperature, and for D we take the value 2·0 for hydrogen atoms diffusing into a mixture of carbon dioxide and water vapour at S.T.P., and use the gas kinetic relation that D is proportional to $T^{\frac{3}{2}}$; this neglects the Sutherland constant, values for which are only known over a small temperature range. The data are set out in the following table:

---

* This is an approximation, neglecting the difference between $\sqrt{(\overline{x^2})}$ and x.

mixture	average temp. assumed, T_r (° K)	value of D	$x = \sqrt{(2Dt_l)}$	δ_l	x/δ_l
C_2H_2-O_2	2200	51·5	$5{\cdot}5 \times 10^{-3}$	$2{\cdot}1 \times 10^{-3}$	2·7
C_2H_2-air	1700	31	$2{\cdot}1 \times 10^{-2}$	$0{\cdot}65 \times 10^{-2}$	3·1
C_2H_2-argon	1900	37	$1{\cdot}4 \times 10^{-2}$	$0{\cdot}45 \times 10^{-2}$	3·1
butane-air	1600	29	$6{\cdot}3 \times 10^{-2}$	$2{\cdot}0 \times 10^{-2}$	3·1

We see that there appears to be a definite proportionality between the thickness of the reaction zone and the distance through which a molecule may diffuse. An atom starting at the end of the reaction zone, where the concentration of free atoms and radicals is highest, moves towards the ignition point by diffusion and is driven back by the flow of gas. According to the $x^2 = 2Dt$ law, the diffusion is initially very fast and slows down as the limit of diffusion is reached. To reach the ignition point an atom will travel by diffusion a distance of approximately $2\delta_r$ in time t_r. This can be shown by differentiating $x = \sqrt{(2Dt)}$ to obtain the velocity, and equating it to the gas velocity at the ignition point. This gives us the proportionality between the distance covered by diffusion and the thickness of the luminous zone which is observed experimentally. Of course, in practice, some individual atoms diffuse more rapidly than others, so that the commencement of chemical action and beginning of luminosity is rather indefinite, as is in fact observed, although the more rapid initial diffusion at the high temperature end of the reaction zone and the slowing down as the atoms come into cooler gas sharpens the limit somewhat. In practice it seems that the distance x travelled by diffusion is very nearly equal to $3\delta_l$. Thus, allowing for the rather big approximations involved in the calculations, and the fact that the thickness of the reaction zone is rather greater than the thickness of the luminous zone, δ_l, measured between contrast zones, it seems that the relation that the distance travelled by diffusion in time t_r is equal to $2\delta_r$ is sufficiently confirmed by experiment. It may therefore be taken as fairly well established that the ignition point, as previously defined by the commencement of marked exothermic reaction, corresponds to the limit to which free atoms or radicals may reach by diffusion against the gas stream.

RELATIONSHIPS BETWEEN FLAME VELOCITY, DIFFUSION, THICKNESS OF THE REACTION ZONE AND GAS PRESSURE

Using the relationship which we have deduced above, that the distance x travelled by diffusion $= 2\delta_r$, we have

$$(2\delta_r)^2 = 2Dt_r,$$

hence the gas velocity in the reaction zone, v_r, is

$$v_r = \delta_r/t_r = D/2\delta_r.$$

Therefore, with sufficient accuracy,

$$v = \frac{D}{2\delta_r} \times \frac{T_i}{T_r}.$$

Remembering equation (1), $vn_c = \bar{u}\delta_r$, we have

$$v^2 = \frac{D \times \bar{u}}{2n_c} \times \frac{T_i}{T_r}. \tag{6}$$

Thus we again arrive at the relationship between the square of the flame velocity and the reaction velocity, but deduced in a more natural way and based on experiment.

From the above relationship (6) we can predict that the flame velocity should be independent of pressure because, n_c is proportional to pressure, D is inversely proportional to pressure, and $\bar{u}$ will be proportional to the square of the pressure if it depends on the total number of two-body collisions occurring per unit volume per sec. This independence of flame velocity on pressure cannot be expected to hold with absolute rigidity because among other factors the amount of dissociation in the flame will vary slightly with pressure and this will affect both the flame temperature (and hence the rate of diffusion) and the concentration of free atoms at the ignition point. Nevertheless we see from (6) that the velocity should not vary very greatly with pressure provided the character of the chemical processes remains similar and the burner diameter is sufficiently great to prevent surface effects from becoming important. In practice (Wolfhard 1943) it has been found that for an oxy-acetylene flame the velocity is quite accurately independent of pressure over the range 1 atm. to 10 mm. and there is no marked change down to 1 mm., and for acetylene-air and butane-air flames the variation is slight. This is in contrast with the calculations of Tanford & Pease which predict that the flame velocity should vary inversely as the fourth root of the pressure. Their calculation assumes that the amount of dissociation is inversely proportional to the square root of the pressure; however, this would only be the case if the flame temperature were independent of pressure. It seems that in practice the flame temperature adjusts itself so that the proportion of free atoms and radicals does not vary so greatly. Thus (see Gaydon & Wolfhard 1948) for an oxy-acetylene flame at 1 atm. pressure the equilibrium mixture contains 12 % of atomic oxygen and 7 % of atomic hydrogen; at $\frac{1}{100}$ atm. the atomic oxygen only rises to $14\frac{1}{2}$ % and that of hydrogen to 13 %, which is very much less than the ten-fold increase given by the inverse square-root law.*

These rather crude calculations, based on the relation between the distance through which a hydrogen atom may diffuse and the observed thickness of the luminous reaction zone, give strong support to the view that the diffusion of free atoms and radicals is of major importance in the mechanism of flame propagation. The prediction that the flame velocity should be nearly independent of pressure, which is confirmed by observation, indicates that the calculations we have made are probably applicable to the hydrocarbon flames studied. The comparatively low temperature at which the ignition point appears to lie, and the very short induction periods required are inconsistent with a purely thermal theory of flame propagation.

---

* In Tanford & Pease's equation for the flame velocity there is a constant, denoted B_i, which they state is independent of pressure; actually it appears to involve the thickness of the reaction zone, which varies with pressure, but even this alteration could not reduce their equation to give a velocity independent of pressure.

We wish to express our thanks to Sir Alfred Egerton for his keen interest in this work, and one of us (A.G.G.) is indebted to the Royal Society for the award of a Warren Research Fellowship.

REFERENCES

Damköhler, G. 1940 *Z. Elektrochem.* **46**, 601.
Egerton, A. C. & Powling, J. 1948 *Proc. Roy. Soc.* A, **193**, 172, 190.
Gaydon, A. G. & Wolfhard, H. G. 1948 *Proc. Roy. Soc.* A, **194**, 169.
Klaukens, H. & Wolfhard, H. G. 1948 *Proc. Roy. Soc.* A, **193**, 512.
Lewis, B. & von Elbe, G. 1934 *J. Chem. Phys.* **2**, 283.
Tanford, C. & Pease, R. N. 1947 *J. Chem. Phys.* **15**, 431, 861.
Van Tiggelen, A. 1946 *Bull. Soc. Chim. Belg.* **55**, 202.
Wolfhard, H. G. 1943 *Z. tech. Phys.* **24**, 206.

The second virial coefficients of organic vapours

BY J. D. LAMBERT, G. A. H. ROBERTS, J. S. ROWLINSON AND V. J. WILKINSON
Physical Chemistry Laboratory, University of Oxford

(*Communicated by Sir Cyril Hinshelwood, F.R.S.—Received* 26 *July* 1948)

The compressibilities of a number of organic vapours have been measured at pressures up to 1 atm. and temperatures ranging from 40 to 130° C. The observed second virial coefficients are compared with values calculated from the critical data by the Berthelot equation. The results show two distinct classes of behaviour. Class I is shown by ethane, ethylene, *n*-hexane, cyclohexane, benzene, diethyl ether, ethyl chloride, chloroform and carbon tetrachloride, where the measured second virial coefficients are in agreement with the calculated values. Class II by acetaldehyde, acetone, acetonitrile, methyl alcohol, where the measured second virial coefficients are consistently very much higher than the calculated values. It is concluded that the vapours of polar substances for which the energy of attraction between molecules, due to dipole interaction or to hydrogen bonding, is larger than kT undergo dimerization. This view is supported by thermal conductivity data. The range of validity of the Berthelot equation for both non-polar and polar vapours is examined.

INTRODUCTION

The Berthelot equation is generally regarded as providing a satisfactory empirical approximation to an equation of state for vapours at pressures below 1 atm. Direct measurements of the compressibility of acetaldehyde vapour (Alexander & Lambert 1941) showed that the second virial coefficients at temperatures between 20 and 100° C were very much higher than the values calculated from critical data by the

Berthelot equation. The discrepancy was interpreted in terms of a partial dimerization of the vapour, and quantitative calculations based on this assumption gave a reasonable value for the heat of dimerization of acetaldehyde molecules by hydrogen bonding. A similar dimerization was subsequently proposed for ethyl chloride (Schäfer & Foz Gazulla 1942) to account for the observed variation of thermal conductivity of the vapour with change of pressure. It therefore seemed interesting to make a series of measurements of the compressibilities of simple organic vapours, both polar and non-polar, in order to discover whether the Berthelot equation of state is generally applicable to non-polar vapours, and whether with polar vapours systematic deviations are found which may be interpreted in terms of dimerization due to dipole interaction or to hydrogen bonding.

EXPERIMENTAL

Method

Measurements were made in a simple 'Boyle's law apparatus' as described by Alexander & Lambert (1941). Temperature was controlled by a constant-pressure vapour jacket and manostat described by Lambert & Clark (1927). By using ethyl alcohol and chlorobenzene as vapour-bath liquids it was possible to obtain steady temperatures over the ranges 40 to 80 and 80 to 130° C. Temperature was measured by a thermometer graduated to 0·2° C suspended inside the vapour jacket. Pressure could be varied from zero to 1·5 atm., and was read by a cathetometer to 0·01 mm. Volume was measured in a 50 ml. burette graduated at 0·1 ml. intervals which could be read to 0·005 ml. using the cathetometer scale. The burette was calibrated by running out and weighing mercury with the burette set in the same position as in the complete apparatus. Burette and manometer limb were both of about 1·0 cm. internal diameter.

Second virial coefficients were obtained from the equation of state

$$pv = n(RT + Bp)$$

by plotting pv against p and measuring the slope and the intercept $p_0 v_0$. Then

$$n = p_0 v_0 / RT \quad \text{and} \quad B = (RT/p_0 v_0) \times (\text{slope of } pv/p \text{ plot}).$$

The accuracy of the apparatus was checked by making measurements with nitrogen at 50° C (the Boyle point). The maximum deviation of the value of pv from the mean was 0·09 % over pressures ranging from 200 to 600 mm. The accuracy of B depends on a number of variable factors affecting the accuracy with which the slope of the pv/p plot can be measured. The most important is the limitation of the range of pressures in which measurements are possible with the less volatile substances. The average expected error is of the order of ± 50 ml./mol., falling to ± 20 ml./mol. for the more volatile substances and rising to ± 100 ml./mol. for the less volatile.

The possibility of errors due to adsorption on the walls of the burette in measurements of this kind was investigated for a number of simple organic vapours by Drucker & Ullmann (1910). They found that only with carboxylic acids was any correction necessary for adsorption. In the present series of experiments additional measurements with a quantity of glass-wool in the burette were made for benzene, ether, methyl alcohol and acetone. There was no measurable difference in the values of B obtained with and without glass-wool, and it was assumed that the effect of adsorption can be neglected.

Materials

(*a*) *Gaseous*

Ethane was prepared by the electrolysis of potassium acetate solution (Klemenc 1938). The gas was passed through towers of pumice soaked with potassium hydroxide solution, to remove carbon dioxide, and with oleum at 70° C, to remove ethylene, into a gas-holder in which it was stored over potash solution. It was then passed through calcium chloride and phosphorus pentoxide tubes into a bulb where it was frozen with liquid air. It was then fractionally distilled *in vacuo* several times and stored in a glass vessel with a mercury sealed tap. Infra-red absorption measurements showed less than 1 part in 8000 of ethylene.

(*b*) *Liquid*

All liquids were frozen by liquid air in a dry bulb sealed to the apparatus and thoroughly pumped out with a mercury-vapour pump to remove air. They were then distilled into a second bulb *in vacuo* (only the middle fraction being condensed) and again frozen and pumped out before use.

n-*Hexane* ('free from aromatic compounds') was shaken with concentrated sulphuric acid followed by alkali and water, dried over phosphorus pentoxide and distilled.

Cyclohexane was shaken with nitric and sulphuric acids to remove aromatic compounds, washed with water, dried over potassium hydroxide pellets followed by phosphorus pentoxide and fractionated twice. Melting-point of purified substance was 5·4° C.

Ethyl chloride ('General Anaesthetic') was squirted directly into a cooled bulb filled with dry air and sealed to the apparatus, and treated as described above. Two different samples gave concordant results.

Carbon tetrachloride ('puriss.') was shaken with alcoholic caustic soda, dried over calcium chloride and fractionated.

Chloroform was washed with distilled water, dried over calcium chloride and fractionated.

Diethyl ether ('Analar') was dried over calcium chloride and fractionated.

Benzene. A sample of very pure benzene which had been specially prepared for vapour-pressure measurements was kindly supplied by Dr D. H. Everett. This had been finally purified by fractional crystallization and stored over sodium wire.

Acetone ('Analar') was dried over potassium carbonate and fractionated.

Methyl alcohol ('Analar') was dried by refluxing with bright magnesium ribbon and fractionated.

Acetonitrile was allowed to stand for 7 days over anhydrous sodium sulphate and for 2 days over phosphorus pentoxide. It was then fractionated.

RESULTS

The measured values of B in ml./mol. are shown plotted against temperature in figures 1 and 2. The results obtained for ethane are plotted together with very accurate measurements made at lower temperatures by Eucken & Parts (1933).

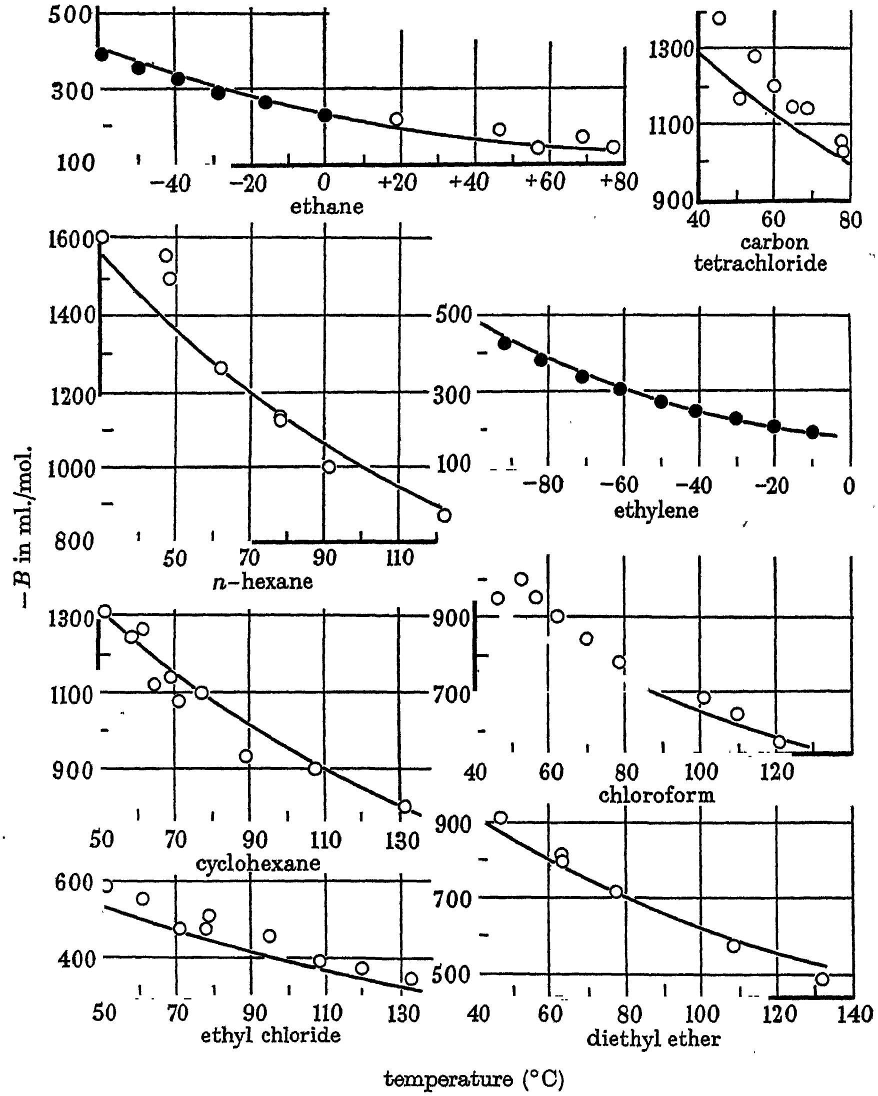


FIGURE 1. ○ Experimental values. ● Results of Eucken & Parts.
—— Curve calculated from Berthelot equation.

Results obtained for ethylene by Eucken & Parts are also shown on a separate graph. It was impossible to make measurements with carbon tetrachloride at temperatures above 80° C, as the vapour attacked the mercury in the burette; the measurements at lower temperatures may be unreliable for the same reason. All plots of pv against

p from which the values of B were obtained were linear. A specimen plot for acetaldehyde was given by Alexander & Lambert (1941); the plots for less volatile compounds showed considerably greater scatter of points, but no definite curvature.

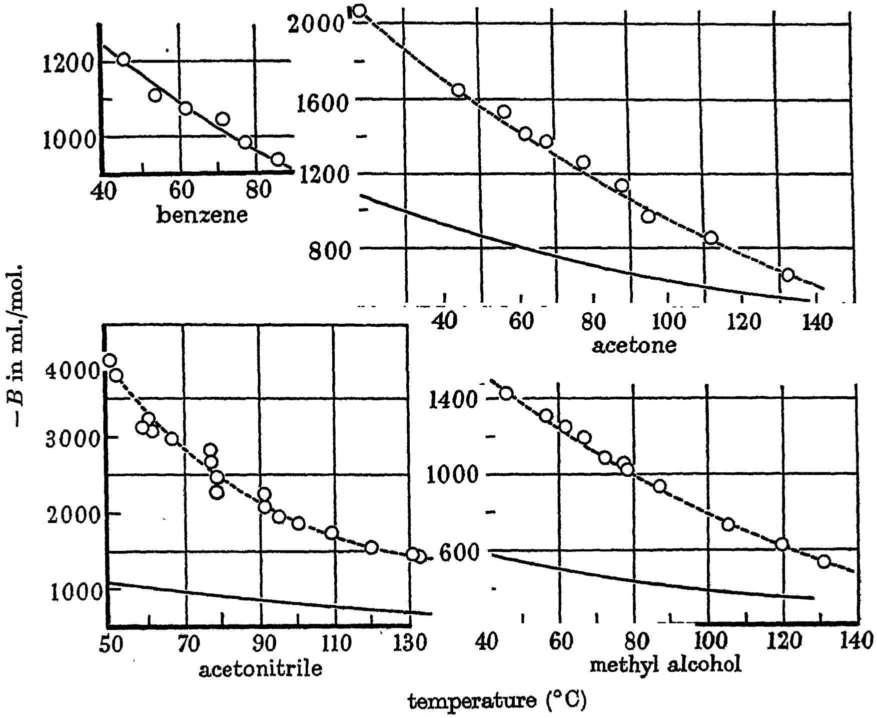


FIGURE 2. ○ Experimental values. —— Berthelot curves.

Values of B calculated from critical data by the Berthelot equation are shown on all the graphs as a full line. Critical data were obtained from Landolt and Bornstein Tables, and B calculated from the equation

$$B = \frac{9}{128}\frac{RT_c}{P_c}\left(1 - \frac{6T_c^2}{T^2}\right).$$

The values used for the various substances were as follows:

Ethane. $T_c = 35{\cdot}9^\circ$ C, $P_c = 48{\cdot}8$ atm., $B = 36{\cdot}5 - 2{\cdot}09 \times 10^7 T^{-2}$.
n-*Hexane.* $T_c = 234^\circ$ C, $P_c = 29{\cdot}5$ atm., $B = 99{\cdot}5 - 15{\cdot}26 \times 10^7 T^{-2}$.
Cyclohexane. $T_c = 281^\circ$ C, $P_c = 40{\cdot}6$ atm., $B = 78{\cdot}8 - 14{\cdot}52 \times 10^7 T^{-2}$.
Benzene. $T_c = 288{\cdot}5^\circ$ C, $P_c = 47{\cdot}7$ atm., $B = 67{\cdot}9 - 12{\cdot}85 \times 10^7 T^{-2}$.
Ethylene. $T_c = 9{\cdot}7^\circ$ C, $P_c = 50{\cdot}9$ atm., $B = 32 - 1{\cdot}537 \times 10^7 T^{-2}$.
Ethyl chloride. $T_c = 183^\circ$ C, $P_c = 54$ atm., $B = 48{\cdot}7 - 6{\cdot}075 \times 10^7 T^{-2}$.
Chloroform. $T_c = 262{\cdot}5^\circ$ C, $P_c = 53{\cdot}8$ atm., $B = 57{\cdot}4 - 9{\cdot}892 \times 10^7 T^{-2}$.
Carbon tetrachloride. $T_c = 283{\cdot}1^\circ$ C, $P_c = 44{\cdot}98$ atm., $B = 71{\cdot}4 - 13{\cdot}25 \times 10^7 T^{-2}$.
Diethyl ether. $T_c = 194{\cdot}6^\circ$ C, $P_c = 36{\cdot}7$ atm., $B = 73{\cdot}5 - 9{\cdot}645 \times 10^7 T^{-2}$.
Acetone. $T_c = 235^\circ$ C, $P_c = 47$ atm., $B = 62{\cdot}4 - 9{\cdot}66 \times 10^7 T^{-2}$.
Methyl alcohol. $T_c = 240^\circ$ C, $P_c = 78{\cdot}6$ atm., $B = 37{\cdot}6 - 5{\cdot}945 \times 10^7 T^{-2}$.
Acetonitrile. $T_c = 274{\cdot}7^\circ$ C, $P_c = 47{\cdot}7$ atm., $B = 66{\cdot}3 - 11{\cdot}93 \times 10^7 T^{-2}$.

The compounds fall clearly into two distinct classes:

Class I. Substances where the measured value of B does not depart significantly from the value calculated by the Berthelot equation: ethane, ethylene, *n*-hexane, cyclohexane, benzene, diethyl ether, ethyl chloride, chloroform, carbon tetrachloride. In the case of the three halogen compounds the measured values of B *all* lie slightly above the calculated curve, but by an amount not significantly greater than the likely experimental error.

Class II. Substances where the measured values of B are consistently very much higher (over 100 % at the lower temperatures) than the calculated values: acetone, methyl alcohol, acetonitrile. Acetaldehyde (Alexander & Lambert 1941) behaved similarly.

DISCUSSION

The nature of the interaction between molecules

It has been pointed out by Eucken (1931) that the effects of attractive forces between molecules of a gas may be twofold in nature. In the first place there will always be an acceleration of molecules as they approach one another, giving rise to the lowering of the observed pressure which is expressed in the various equations of state. In the second place, if a triple collision occurs in which the third partner can take away the energy of approach, a 'double molecule' of low stability may be formed. In the case of the fairly complex organic molecules here considered, each with several vibrational degrees of freedom, such a double molecule might be formed in a binary collision. Its stability will, of course, depend on the magnitude of the energy of attraction between the two molecules. Of the compounds whose second virial coefficients have been measured the more highly polar, which might be expected to show strong intermolecular attraction, all fall into class II (abnormally high value of B); the non-polar and less highly polar compounds fall into class I (Berthelot value of B). It therefore seems reasonable to suppose that a dimerization occurs with the molecules in class II, giving rise to the abnormally high second virial coefficient, while, for the substances in class I, the intermolecular forces are insufficient for the formation of a dimer, and only give rise to the general acceleration of molecules approaching one another.

Reversible dimerization in a vapour would be expected to cause an abnormal increase in thermal conductivity with rise of pressure. Such an increase was demonstrated by Schäfer & Foz Gazulla (1942) for ethyl chloride. Preliminary measurements made in this laboratory by Mr S. D. Woods have shown that for the vapour of diethyl ether at 25° C there is no measurable change in thermal conductivity between 100 and 400 mm. pressure, while for the vapour of acetaldehyde there is a 2·5 % rise in thermal conductivity between these pressures (confirming earlier measurements by Milverton (1935) on acetaldehyde). Mann & Dickins (1932) measured the thermal conductivity of ethane over a range of pressures and found no variation greater than could be attributed to convection under their experimental conditions. Professor Foz Gazulla (private communication) has recently made measurements with benzene vapour, which shows no measurable change in conductivity with pressure, and with methyl alcohol, which shows a large increase. These results are in agreement with the view that class II vapours, acetaldehyde, methyl alcohol, undergo reversible dimerization, while class I vapours, ethane, benzene, diethyl ether, do not. The one anomalous case is ethyl chloride which is discussed in detail later.

Validity of the Berthelot equation

The Berthelot equation appears to be valid for a variety of vapours over a large range of temperatures. The molecular model from which it was derived is recognized to be inadequate, which would suggest a much greater limitation in applicability of the equation than has actually been found. It is therefore interesting to examine the reasons for its success on more fundamental theoretical grounds.

The problem of intermolecular forces for gases composed of spherical molecules has been treated by Lennard-Jones (1924) and is summarized by Fowler & Guggenheim (1939). E_r, the energy of interaction of two molecules separated by a distance r, may be represented by the equation

$$E_r = \frac{\nu}{r^{12}} - \frac{\mu}{r^6},$$

where ν and μ are constants. The value of the second virial coefficient B may then be expressed by the following equations:

$$B = \tfrac{2}{3}\pi d^3 N\,.\,F(y),$$

$$y = 2\left(\frac{-E_0}{RT}\right)^{\frac{1}{2}},$$

$$F(y) = y^{\frac{1}{2}}\left[\Gamma(3/4) - \frac{1}{4}\sum_{s=1}^{\infty}\Gamma\left(\frac{2s-1}{4}\right)\frac{y^s}{s!}\right],$$

where E_0 is the maximum energy of interaction and occurs when $r = r_0$. d is the closest distance of approach for slow molecules and is given by the equation

$$r_0 = d\,.\,2^{\frac{1}{6}}.$$

From these equations it follows that a plot of $\log(-F(y))$ against $-2\log y$ can be superposed on a plot of $\log(-B)$ against $\log T$. Corner (1948) has shown that equations of similar form may be applied to cylindrical molecules.

The critical temperature of a gas is a function of E_0; to a very rough approximation $RT_c/-E_0 = 1\cdot3$ for several gases (Lennard-Jones & Devonshire 1937). The value chosen for y is thus a function of T_c/T for the substance under consideration at a temperature T. From the point of view of intermolecular forces the more interesting temperature region is above the critical temperature and few values of $F(y)$ have been published for values of y much greater than 2, corresponding to $T_c/T \leqslant 1\cdot3$. The measurements described here extend over a lower range of temperatures giving T_c/T between 0·8 and 2·1. A fresh computation has therefore been made of $F(y)$ over the range $y = 1\cdot1$ to $y = 4\cdot0$. The expansion was summed to twenty terms giving an accuracy of 1/1000. The resultant plot of $\log(-F(y))$ against $-2\log y$ is shown in figure 3. It will be seen that between the limits $2\log y = 0{\cdot}3$ and $2\log y = 1{\cdot}1$, corresponding roughly to $T_c/T = 0{\cdot}6$ and $T_c/T = 4{\cdot}0$, the points fall very close to a straight line of slope -2. Since this curve is superposable on the plot of $\log(-B)$ against $\log T$, the virial coefficient B should be approximately represented between these limits by an equation containing as its chief term an inverse second power of the temperature. The Berthelot equation, $B = a - bT^{-2}$, fulfils this requirement, since, in the range of temperatures considered, $a \ll bT^{-2}$, and the values of the constants a and b calculated from the critical data seem to give a very fair empirical fit with measured values of B.

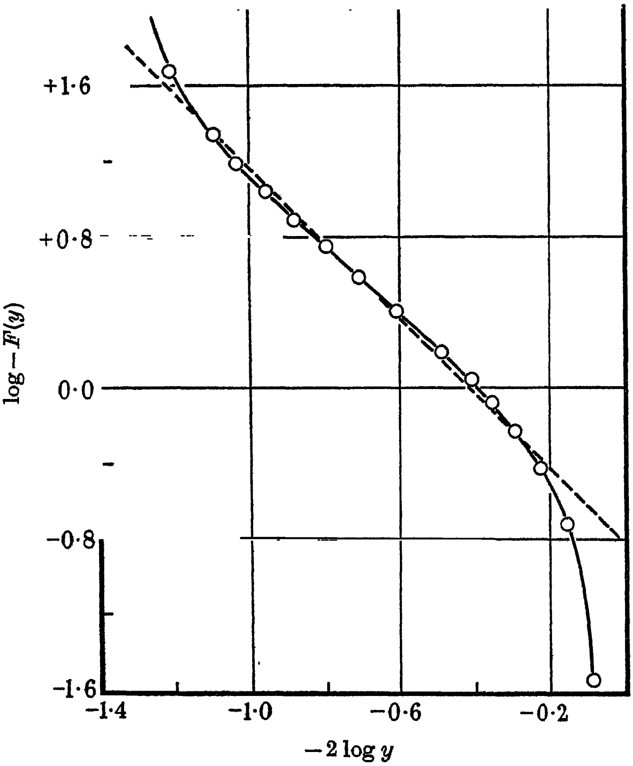


Figure 3. ○ Computed points. ---- Slope of −2.

The above calculations would be expected to apply to *non-polar* substances where the attraction between molecules is due to London dispersion forces for which $E_r = -m/r^6$. Since the Berthelot equation also applies with considerable accuracy to some *polar* molecules, e.g. diethyl ether ($\mu = 1{\cdot}15$), it would appear that the equation, with constants obtained from the critical data, also gives an adequate representation of the effects of dipole interaction, in so far as these are not sufficiently powerful to give rise to the formation of double molecules. The effect of this kind of 'general' dipole interaction is probably small compared with that of the dispersion forces, and furthermore obeys an equation of form $E_r = \frac{-2\mu^4}{3r^6}\frac{1}{kT}$ (London 1937), which shows the same dependence on r as the dispersion forces. It is accordingly assumed in the following section of this discussion that, for those polar molecules where dimerization occurs, the whole of the effect of 'generalized' intermolecular attraction (as opposed to dimer formation) can be represented by the Berthelot equation. The effect of the dimerization is then regarded as superimposed on this. The experimental curves showing B against T (figures 1 and 2) all approach the Berthelot curve with rising temperature indicating that the dimerization becomes negligible before the critical temperature is reached. It is thus reasonable to use the observed critical data for calculating the Berthelot constants, even for substances which undergo dimerization.

Dimerization

The effect of dimerization on the observed second virial coefficient of a vapour has been elegantly treated by Schäfer & Foz Gazulla (1942). Consider one mol. of a gas A of which a small fraction α is associated to A_2:

$$\underset{(1-\alpha)}{2A} \rightleftharpoons \underset{\frac{1}{2}\alpha}{A_2} \quad \text{mols.}$$

If the total pressure is p, then

$$K_p = \frac{p_A^2}{p_{A_2}} = \frac{2p(1-\alpha)^2}{\alpha(1-\frac{1}{2}\alpha)}.$$

For small values of α, where $1 \gg \alpha$ and terms in α^2 may be neglected,

$$\frac{\alpha}{2} = \frac{p}{K_p}.$$

The second virial coefficient of a vapour is defined by the equation

$$pv = n(RT + Bp).$$

If B is the second virial coefficient of the monomer A, the equation of state for a mixture of n_1 mols of A with a very small number n_2 mols of A_2 may be written (to the first order of small quantities)

$$\frac{pv}{RT} = (n_1 + n_2) + \frac{Bn_1p}{RT},$$

$$(n_1 + n_2) = (1 - \alpha/2) = \left(1 - \frac{p}{K_p}\right) \quad \text{and} \quad n_1 \doteqdot 1,$$

so

$$pV = RT + p\left(B - \frac{RT}{K_p}\right).$$

The vapour thus follows an equation of state

$$pV = RT + B_0p,$$

where the observed second virial coefficient

$$B_0 = \left(B - \frac{RT}{K_p}\right).$$

It is possible to calculate K_p from the measured second virial coefficient B_0 if it is assumed that the Berthelot equation gives the 'true' value of B for the monomeric vapour. Values of K_p have been calculated from the measurements on acetone, methyl alcohol and acetonitrile, and the plots of $\log_{10} K_p$ against $1/T$ are shown in figure 4. The values for acetaldehyde (Alexander & Lambert 1941) have been recalculated over the whole temperature range by this improved method and are also plotted in the figure. The points for acetaldehyde and acetonitrile fall on straight lines whose slopes correspond to heats of dimerization $\Delta H = -4600$ cal./mol. for acetaldehyde, and $\Delta H = -5200$ cal./mol. for acetonitrile. The points for acetone and methyl alcohol fall on curves whose slope increases with rise of temperature. For methyl alcohol the apparent value of $-\Delta H$ varies between 3200 and 7300 cal./mol., and for acetone between 3200 and 8010 cal./mol.

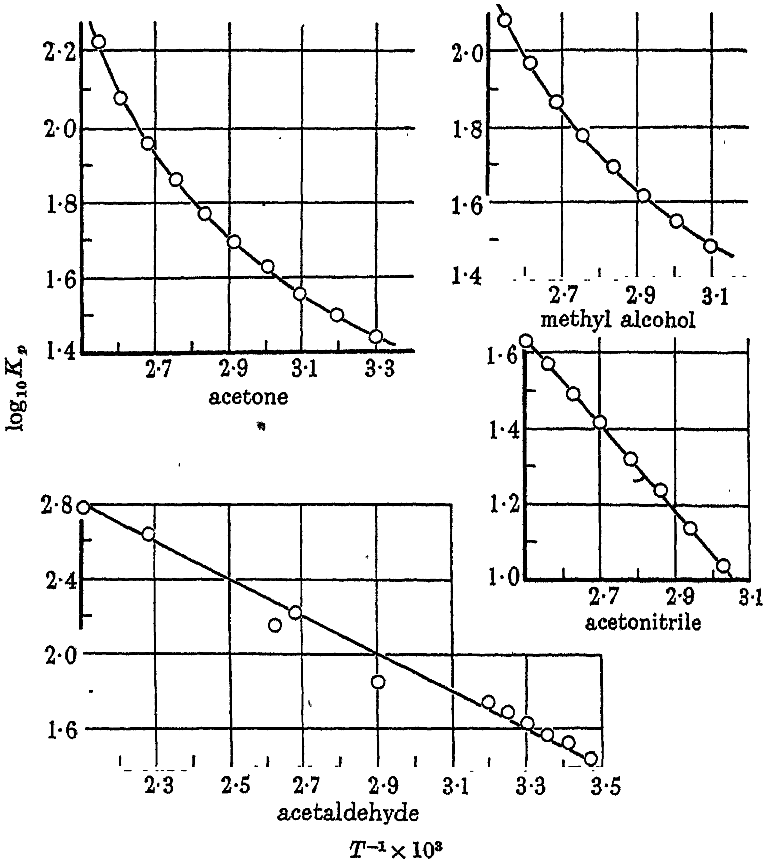
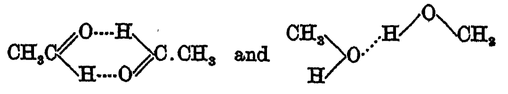


Figure 4

The forces responsible for dimerization could be either dipole interaction or hydrogen bonding. The latter would be expected only with acetaldehyde and methyl alcohol, which could form dimers of structure



The accepted values for the bond energy of O—H---O range between 4500 and 8200 cal./mol. (Pauling 1940), which is in fair agreement with the heat of formation of the methyl alcohol dimer. No value has been estimated for a C—H---O bond. The structure suggested above for the acetaldehyde dimer would correspond to a bond energy of 2300 cal./mol. per hydrogen bond. This compares reasonably with the values given (Pauling 1940) for C—H---N (3300 to 4300 cal./mol.) and for N—H---N (1300 cal./mol.).

Dipole interaction can, of course, occur with all polar molecules. Values of the energy of dipole interaction have been calculated using the approximation formulae: $\Delta U = -\mu^2/r^3$ for parallel pairs of dipoles, and $\Delta U = -2\mu^2/r^3$ for pairs of dipoles arranged end to end, where μ is the dipole moment and r the closest distance of approach of the centres of the dipoles. r was obtained by taking the covalent and the van der Waals radii of atoms and the bond angles given by Pauling (1940) and estimating the nearest distance of approach of the molecular dipoles, regarding the molecules as combinations of rigid van der Waals radii. The results are given in table 1. The values of $-\Delta U$ are minimum values, since, if two molecules form a dimer owing to dipole attraction, they probably lie closer together than the van der Waals radii would indicate. The calculated energies of dipole interaction are thus of the right order to account for the observed heats of dimerization of all four substances, acetaldehyde, acetonitrile, acetone and methyl alcohol. It is to be expected that stable dimers will only be formed when $-\Delta U \gg RT$. This requirement is fulfilled for the above four substances, where the values of $-\Delta U/RT$ range between 2·8 and 8·3 under the conditions of experiment. It is not fulfilled for diethyl ether and chloroform, where no evidence of dimerization was found, and for which $-\Delta U/RT \leqslant 0{\cdot}9$. Ethyl chloride, where $-\Delta U/RT$ lies between 1·7 and 2·2, may be regarded as a borderline case, and may show a degree of dimerization sufficient to influence the thermal conductivity measurements, but not the less accurate virial coefficient measurements.

TABLE 1

substance	$\mu \times 10^{18}$ (e.s.u.)	r (Å)	$-\Delta U$ calc. (cal./mol.)	$-\Delta H$ obs. (cal./mol.)	$-\Delta U/RT$ 40° C	$-\Delta U/RT$ 130° C
acetonitrile	3·5	3·1 $\uparrow\downarrow$	6280	5200	8·3	6·5
acetone	2·74	3·2 $\uparrow\downarrow$	3320	3200–8010	5·3	4·1
acetaldehyde	2·7	3·2 $\uparrow\downarrow$	3380	4600	5·4	4·2
methyl alcohol	1·66	2·6 $\uparrow\downarrow$	2270	3200–7300	3·6	2·8
ethyl chloride	2·02	3·5 $\uparrow\downarrow$	1370	—	2·2	1·7
chloroform	1·1	4·0 $\overset{\uparrow}{\uparrow}$	550	—	0·9	0·7
diethyl ether	1·15	4·0 $\overset{\uparrow}{\uparrow}$	600	—	0·9	0·7

$\uparrow\downarrow$ indicates that the maximum energy of interaction is found when the dipoles are parallel to one another.

$\overset{\uparrow}{\uparrow}$ indicates that the maximum energy of interaction is found when the dipoles are arranged end to end.

It is difficult to account for the apparent increase in heat of dimerization of methyl alcohol and acetone with rise of temperature, as contrasted with the constant heat of dimerization found with acetonitrile and acetaldehyde. Acetonitrile is a linear molecule and might be expected to behave like a rigid dipole. The other three molecules are non-linear, and it may be that the increased intramolecular vibrations at higher temperatures enable the dipoles of methyl alcohol and acetone to come closer together in the initial collision. If, on the other hand, the dimerizations of acetaldehyde and methyl alcohol are due to hydrogen bond formation, both would be expected to show similar behaviour. It may be that formation of higher polymers occurs with methyl alcohol and acetone, which would invalidate the quantitative calculations. This would, however, be expected to give a curved pv/p plot, which was not in fact observed with these vapours.

CONCLUSION

The general conclusion is that the attractive forces between molecules of a vapour produce two distinct effects:

(1) A general acceleration of molecules as they approach one another, resulting in a diminution of the observed pressure. This occurs with all substances and can be quantitatively expressed for pressures up to one atmosphere and temperatures between $0{\cdot}5\,T_c$ and $1{\cdot}2\,T_c$ by a second virial coefficient calculated from the critical data by the Berthelot equation. This applies to both polar and non-polar molecules.

(2) A formation of double molecules held together by dipole interaction or by hydrogen bonds. This occurs only with polar substances where the energy of interaction between the molecules is larger than kT. Its effect on the compressibility of the vapour is superimposed on that of 1.

It has been shown by Foz Gazulla & Vidal (1947) that published data for the compressibilities of steam and of ammonia may be interpreted in terms of a similar dimerization and that reasonable values for the heats of dimerization are found. An alternative treatment of these vapours was given by Stockmayer (1941), who included the whole effect of dipole interaction in a calculation by Lennard-Jones's method. This gave a fairly successful fit with experimental values, but has the disadvantage of involving many rather arbitrary assumptions about the values of constants and the exact molecular diameter. Measurements of the thermal conductivity of steam (Milverton 1935) show a marked increase with rise of pressure, which is in accord with the view that dimerization occurs.

The authors are grateful to Dr B. Lambert for much help in the design and construction of the apparatus. The measurements on n-hexane were made by Mr R. G. Vines.

REFERENCES

Alexander, E. A. & Lambert, J. D. 1941 *Trans. Faraday Soc.* 37, 421.
Corner, J. 1948 *Proc. Roy. Soc.* A, **192**, 275.
Drucker, K. & Ullmann, G. 1910 *Z. phys. Chem.* 74, 567.
Eucken, A. 1931 *Z. phys. Chem.* (Bodenstein Festband), p. 432.
Eucken, A. & Parts, A. 1933 *Z. phys. Chem.* B, 20, 184.
Fowler, R. H. & Guggenheim, E. A. 1939 *Statistical thermodynamics.* Cambridge University Press.
Foz Gazulla, O. R. & Vidal, J. M. 1947 *An. Soc. esp. Fis. Quim.* 43, 842.

Klemenc, A. 1938 *Die Behandlung und Reindarstellung von Gasen.* Leipzig: Akademische Verlagsgesellschaft.
Lambert, B. & Clark, A. M. 1927 *Proc. Roy. Soc.* A, **117**, 183.
Lennard-Jones, J. E. 1924 *Proc. Roy. Soc.* A, **106**, 463.
Lennard-Jones, J. E. & Devonshire, A. F. 1937 *Proc. Roy. Soc.* A, **163**, 53.
London, F. 1937 *Trans. Faraday Soc.* **33**, 8.
Mann, W. B. & Dickins, B. G. 1932 *Proc. Roy. Soc.* A, **134**, 77.
Milverton, S. W. 1935 *Proc. Roy. Soc.* A, **150**, 287.
Pauling, L. 1940 *Nature of the chemical bond*, 2nd ed. Cornell University Press.
Schäfer, K. & Foz Gazulla, O. R. 1942 *Z. phys. Chem.* B, **52**, 299.
Stockmayer, W. H. 1941 *J. Chem. Phys.* **9**, 398.

The structure and properties of the alloy Cu_2MnIn

By B. R. Coles, W. Hume-Rothery, F.R.S. and H. P. Myers

(*Received* 26 *July* 1948)

The ferromagnetic Heusler alloy Cu_2MnAl has a body-centred cubic structure with an ordered arrangement of the atoms. Considerations of valency and atomic diameters suggested that if the aluminium were replaced by indium the superlattice would be more stable, and that a homogeneous alloy would be obtained if the indium content were slightly less than that required by the formula Cu_2MnIn. An alloy of composition $Cu_{2\cdot057}Mn_{0\cdot975}In_{0\cdot965}$ was prepared and was ferromagnetic in the as-cast state, and also after a variety of heat treatments. The intensities of magnetization were measured in field strengths between 7250 and 17,150 oersteds at temperatures between $-183°$ C and the Curie point. X-ray powder photographs show that the indium alloy has the same structure as the aluminium Heusler alloy but with a larger lattice spacing. This increase in the size of the unit cell means an increase in the R/r value for the closest approach of manganese atoms from 2·84 in Cu_2MnAl to 2·98 in Cu_2MnIn, and corresponds with a fall in Curie point of 97° C. The two alloys are strictly analogous as regards structure and valencies, and the results suggest that in the indium alloy the critical value of R/r is being approached at which ferromagnetism will disappear. Reasons are given for supposing that the ferromagnetism is due to interaction between manganese atoms, in which case the Bohr magneton number per atom of manganese is approximately 4, and manganese behaves as a univalent element, in agreement with conclusions reached independently from the study of phase-boundaries in ternary Cu-Mn-Al alloys.

1. INTRODUCTION

The well known Heusler alloys have a composition given approximately by the formula Cu_2MnAl. An alloy of this composition on slow cooling is non-magnetic and has a structure resembling that of γ-brass, whilst the ferromagnetic form obtained by quenching has an ordered body-centred cubic lattice. The structure of the ferromagnetic alloy was deduced independently by Bradley & Rodgers (1936) and by Heusler (1934) and is shown in figure 1.

The unit cell contains eight of the small body-centred cubic cells, and there are four kinds of atomic position which may be denoted *a*, *b*, *c* and *d*. In the ideal superlattice structure the manganese and aluminium atoms occupy positions *d* and *b* respectively, whilst the *a* and *c* positions are occupied by copper atoms. According to Heusler (1934) the quenched alloy is not perfectly ordered, and on ageing at approximately 100° C the degree of order is increased, and the magnetic properties are improved. The previous work suggested, therefore, that the ferromagnetic properties of an alloy of this type depend on placing the atoms of manganese in a body-centred cubic structure with the further condition that a superlattice must be formed in which the manganese atoms are prevented from being either closest or second closest neighbours. It has been shown by Hume-Rothery (1948) that in ternary alloys of manganese with copper and zinc or aluminium, the phase boundaries in the equilibrium diagram suggest that manganese acts as a divalent element in copper-rich alloys, and as a univalent element in more concentrated alloys. If manganese be univalent, the composition Cu_2MnAl corresponds to an electron/atom ratio of 3/2 and so satisfies the condition for the formation of a body-centred cubic structure. In this alloy it is clear that the superlattice is not very stable. The arguments of Hume-Rothery & Powell (1935) suggested to us that the stability of the superlattice would be increased if either the manganese or the aluminium were replaced by an element of larger atomic diameter. As the ferromagnetic properties depend on the manganese, it was the aluminium which had to be changed for an element of the same valency but of somewhat larger atomic diameter, and for this purpose indium appeared to be suitable. The work of Hume-Rothery, Reynolds & Raynor (1940) has shown that in the binary system copper-indium a typical body-centred cubic β-phase is formed although its composition lies entirely on the copper-rich side of that corresponding to the formula Cu_3In. This effect was regarded as due to the existence of short-range order in the alloy, and suggested that for the present purpose an alloy should be prepared with an indium content very slightly less than that of the formula Cu_2MnIn. The present paper describes the preparation and properties of this alloy and shows that the predictions have been confirmed.

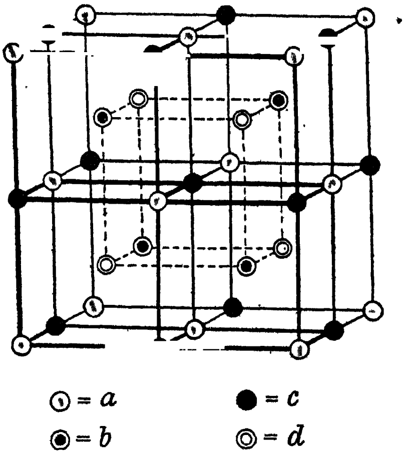


FIGURE 1

2. PREPARATION OF ALLOYS

The alloys were prepared from some especially pure copper kindly presented by the British Non-Ferrous Metals Research Association, indium metal of the highest purity made by Messrs Johnson Matthey and Co. Ltd., and pure electrolytic manganese which was treated in hydrogen in order to remove surface oxidation. Alloys were first prepared by melting the three metals together in an induction furnace, using an atmosphere of hydrogen. It was later found more convenient to prepare first a master alloy of copper and manganese which was remelted with the required amount of indium, using a resistance furnace and a flux of potassium chloride. The resulting alloys were obtained in the furnace-cooled state or were cast into graphite moulds to give ingots $\frac{5}{8}$ in. in diameter. All alloys were given a preliminary homogenization treatment of 4 days at 534° C, and were then treated in different ways as described below. The annealing experiments were carried out in sealed, evacuated glass tubes. The alloy used for the magnetic measurements was analyzed after the heat treatments B, C and D referred to on p. 129. All three metals were determined and the percentage totals lay between the limits 99·94 and 99·97. These totals refer to material which had been annealed and quenched in water; it was found that the quenching produced slight superficial oxidation, and from the above figures it is clear that the melting and annealing techniques gave alloys of high purity.

3. CRYSTAL STRUCTURE AND LATTICE SPACING

Several alloys were prepared with compositions near to that of Cu_2MnIn. All were ferromagnetic in the cast state and after different heat treatments, but microscopic examination showed that some were two-phase in structure. Attention was finally concentrated on an alloy whose composition corresponded to the formula $Cu_{2\cdot057}Mn_{0\cdot975}In_{0\cdot965}$ (44·4 % copper, 18·2 % magnesium and 37·4 % indium by weight). Under the microscope this alloy was homogeneous after quenching from 534° C, and after cooling in air from this temperature, and also after very slow cooling from 534° C, although as shown later (p. 129) the magnetic measurements suggest that very slight sub-microscopic precipitation of a non-magnetic phase may have occurred during the slow cooling.

Debye-Scherrer diffraction films were taken, using unfiltered iron radiation.* Alloys which had been annealed were brittle and could be ground to powder in a mortar, but the resulting Debye-Scherrer films showed fuzzy lines indicating that the material is able to undergo plastic deformation. The powder patterns from filings which were annealed and slowly cooled were indistinguishable from those of filings quenched from 534° C. Both showed the typical body-centred cubic lines together with the superlattice lines expected for the Heusler type of structure. It is thus clear that, in agreement with expectation, the superlattice of Cu_2MnIn is much more stable than that of Cu_2MnAl.

As shown by Bradley & Rodgers (1936), the Debye-Scherrer lines of the Heusler structure fall into three groups: (1) the lines of the 220, 400, 440, ... reflexions which are those of the simple body-centred cubic structure with the indices doubled so as to refer to the large unit cell of figure 1 instead of to the small cubes; (2) the superlattice lines of the 111, 311, ... reflexions with h, k and l odd; and (3) the superlattice lines of the 200, 420, ... reflexions with h, k and l even. In the Heusler alloy Cu_2MnAl the aluminium atom has the lowest scattering power of the three, and the superlattice lines of class (2) are in general weaker than those of class (3). In the alloy Cu_2MnIn the indium atom has the highest scattering power, and it is easy to show that this results in the superlattice lines of class (3) being on the whole stronger than those of class (2). Table 1 shows the relative intensities of the lines calculated for iron $K\alpha$ radiation taking into account the abnormal scattering power in the region of the absorption edge, together with the visual estimates of intensities. It will be seen that the expected reversal of the intensities of superlattice lines of odd and even indices is confirmed. In the paper of Bradley & Rodgers (1936) calculations were also made of the relative intensities of Debye-Scherrer lines from structures in which aluminium atoms occupied the *b* positions in figure 1, and copper and manganese atoms were assumed to occupy the *a*, *c* and *d* positions in ways other than that of the Heusler alloy structure. We have made similar calculations for the alloy Cu_2MnIn and the resulting intensities of the superlattice lines cannot be reconciled with those observed. It is thus clear that the alloy Cu_2MnIn has the true Heusler structure with the aluminium atoms in the *b* positions, the manganese in the *d* positions, and the copper atoms in the *a* and *c* positions, and that the superlattice is extremely stable.

---

* The experimental methods on the X-ray side were those which are now standard in this laboratory and have been described on many occasions, cf. Axon & Hume-Rothery (1948).

TABLE 1

line	calculated intensity (Fe $K\alpha$)	visual estimate
111	2·2	faint
200	1·0	v.v. faint
311	7·9	medium
331	5·9	medium
420	1·8	v. faint
511 333	11·8	medium
531	26·5	strong
600	1·2	v.v. faint

In this table the second column gives the relative calculated intensities of the superlattice lines of Cu_2MnIn corrected for both absorption and temperature, the intensity of the (200) line being taken as unity. The temperature correction was made by assuming a characteristic temperature $\theta_D = 240$ which was obtained by using the melting-point equation of Lindemann; the alloy was found to melt at approximately 630° C. It will be appreciated that, although an error in the temperature factor will affect the relative intensities of lines of high and low angle, it will have little effect on adjacent lines, and it is the comparison of adjacent lines which enables one to reject conclusively the alternative structures discussed by Bradley & Rodgers. The intensities of the main lines were all strong, in agreement with expectation.

The lattice spacing of the alloy $Cu_{2\cdot057}Mn_{0\cdot975}In_{0\cdot965}$ was determined as $a = 6\cdot1865$ Å for quenched specimen 2B at 19·5° C, and as $a = 6\cdot1881$ Å for air-cooled specimen 2C at 19° C.

4. MAGNETIC PROPERTIES

The saturation intensities of the alloy specimens were measured by the ring balance evolved by Sucksmith (1939). The alloys were too brittle to be worked to shape, and were therefore examined as coarse crushed powder* which was packed into the cylindrical specimen container 4 mm. long by 2 mm. diameter. The specimens weighed about 40 mg. and as the intensity of magnetization of the alloy was low, the effect of the demagnetizing factor was readily overcome in the relatively large magnetic fields used (7000 to 17,000 oersteds).

---

* The diameter of the particles was of the order $\frac{1}{2}$ mm.

The saturation intensity of magnetization per unit mass for each specimen was measured at −183° C, −75° C, room temperature, and at temperatures up to Curie point. All the measurements were made relative to pure iron (Johnson and Matthey) as standard, the intensity per unit mass being taken as 217·8 at 17° C (Weiss & Forrer 1929). The mass of the pure iron specimen was adjusted so that it gave deflexions of the same order as those of the alloy specimen.

The alloy of composition $Cu_{2\cdot057}Mn_{0\cdot975}In_{0\cdot965}$ was examined after the following heat treatments:

A. Homogenized for 4 days at 534° C, quenched and crushed.

B. Treated similarly to A, and then reannealed for 12 hr. at 534° C and quenched.

C. Treated similarly to A and then reannealed for 12 hr. at 534° C and allowed to cool in air.

D. Similar to C but cooled very slowly.

The intensities of magnetization per unit mass were measured in external magnetic fields of seven strengths between 7250 and 17,150. The field inside the specimen could not be calculated because the demagnetizing factor for the powder was unknown.* Extrapolation to infinite field strength was thus not possible, but as the alloy saturates easily in the field used, little error results from taking the value of the intensity in the highest field strength (17,150) instead of that in infinite field strength.

The saturation intensities at room and low temperatures are shown in table 2 from which it will be seen that the quenched specimen B has the highest magnetization. The slightly lower values for specimen D may be the result of submicroscopic precipitation of a non-magnetic phase during the very slow cooling; this hypothesis would agree with the fact that the figures for the more rapidly cooled specimen C are almost the same as those for specimen B. The material of specimen A (crushed after homogenization but not reannealed) differs only slightly from those of specimens B, C and D, and the figures in table 2 show that when once the material has been homogenized, its magnetic properties are comparatively little affected by later treatments, either mechanical or thermal.

In figure 2 the values of σ for specimen B at 290, 197 and 90° K are plotted as functions of T and T^2, and it will be seen that the T^2 relation is approximately a straight line in agreement with the general conclusions of Fallot (1935). We believe that the straight line law does not hold exactly at temperatures as high as 291° K, and have therefore extrapolated the straight line through the two lowest points in order to obtain the saturation value σ at the absolute zero. The value of σ_0 for the external magnetic field $H = 17,150$ oersteds is 74·4 ergs/oersted/g. On the assumption that the magnetic properties of the alloy are due entirely to the manganese, this

---

* If, however, the powder is assumed to act as a collection of independent spherical grains the value of the demagnetizing field for a particular specific intensity of magnetization σ, would be $\frac{1}{3}(4\pi\rho\sigma)$, which for $\rho = 8$ and $\sigma = 64$ is 2100 oersteds.

TABLE 2. SATURATION INTENSITY OF MAGNETIZATION PER UNIT MASS AT ROOM AND LOW TEMPERATURES

speci-men	temperature (° C)	external field 7250	8880	10,800	12,720	13,740	15,680	17,150
2A	18	62·5	62·8	62·9	63·0	63·1	63·3	63·3
	− 75	67·3	67·6	67·9	68·1	68·3	68·4	68·4
	−183	71·1	71·5	72·0	72·1	72·3	72·7	72·6
2B	18	62·8	63	63·1	63·3	63·3	63·3	63·5
	− 75	68·4	68·5	68·9	68·9	68·9	69	69·1
	−183	72·4	72·4	72·8	72·9	73	73·1	73·3
2C	18	63	63·1	63·4	63·5	63·6	63·7	63·8
	− 75	68·2	68·5	68·8	68·8	68·8	69	69
	−183	72	72·3	72·7	72·6	72·6	72·8	72·9
2D	18	62·6	62·6	62·8	63	63·1	63·2	63·5
	75	67·9	68·3	68·4	68·4	68·5	68·5	68·7
	−183	71·6	72·1	72·3	72·2	72·3	72·4	72·6

Curie points

H_s	Curie point (° C) 2A	2B	2C
7,250	240	232	212
17,150	250	240	222

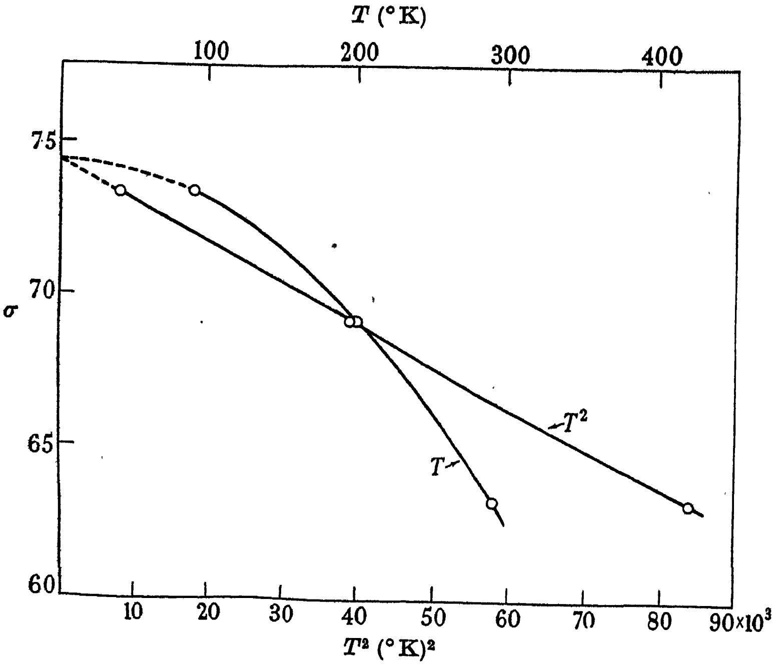


FIGURE 2. $H_\epsilon = 17{,}150$, $\sigma_0 = 74\cdot4$, $p_B = 4\cdot04(6)$.

corresponds to a Bohr magneton value for the manganese atom of 4·04(6).* The validity of this assumption is discussed later.

The high temperature measurements were carried out *in vacuo* and the variation of σ of specimen B with temperature is shown in figure 3, indicating an apparent Curie temperature of 243° C in an external field of 17,150 oersteds, and of 233° C in a field of 7250 oersteds. The true Curie temperature, that appertaining to zero external field strength, would not differ greatly from the latter value.

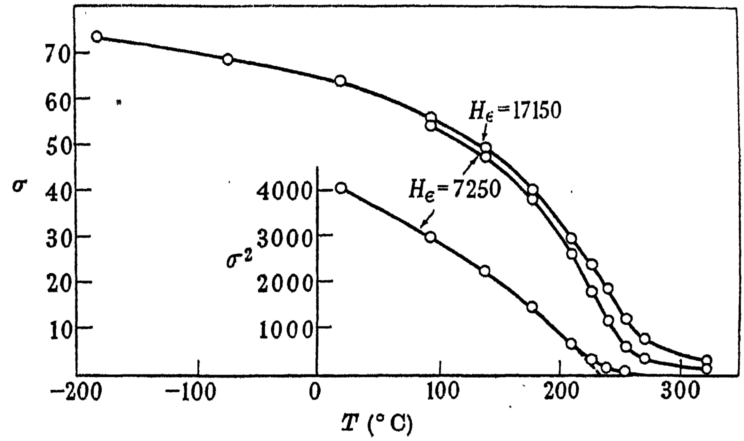


FIGURE 3

DISCUSSION

In table 3 we show some physical properties of the ferromagnetic alloys Cu_2MnAl, taking the Bohr magneton value and Curie point obtained by Valentiner & Becker (1933), Cu_2MnIn, and of the analogous alloy Cu_2MnSn examined by Carapella & Hultgren (1941). The body-centred cubic form of this last alloy is stable only at high temperature, and the data refer to quenched material. If manganese and tin are assumed to be univalent and tetravalent respectively, the composition Cu_2MnSn no longer corresponds to an electron concentration of 3/2, but in the Cu-Mn-Sn system the β-phase area at high temperature is sufficiently wide to include this composition. The three alloys form a well-defined group in which the Bohr magneton number of the manganese atom is approximately 4, if the whole of the ferromagnetism is associated with the manganese atoms; if this assumption is invalid a Bohr magneton number of approximately 4N is associated with an atomic group of the type $(Cu_2MnX)N$, where X = Al, In or Sn, and N expresses the unknown molecular weight of the grouping responsible for ferromagnetism.

According to the accepted theory, ferromagnetism requires the presence of atoms with an incomplete shell of electrons to provide a permanent moment, and in practice only *d* and *f* shells need be considered because the radius of the incomplete shell must be large. The atoms with incomplete shells must have neighbours with which they can interact to produce a positive value of the exchange integral, and hence the distance between the atoms must not be too great, or else the interaction energy becomes insufficient to stabilize the ferromagnetic condition. Slater has shown that in general positive interaction occurs when the ratio of the atomic diameter (i.e. the internuclear distance) to the 'diameter' of the incomplete electron shell is greater than about 1·5. Stoner (1930) has also shown in his simplified interchange interaction theory of ferromagnetism that the following relationship is in general terms correct:

---

* In the calculation of this quantity the value of the Bohr magneton was taken to be 5564 units/g. atom. The recent revision of the fundamental constants suggests that the value 5580 units/g. atom is more probable, but we have retained the older value in order that comparison may be made with other work.

$$J = \frac{\mathrm{k}\theta}{Z},$$

TABLE 3

	Cu_2MnIn	Cu_2MnAl	Cu_2MnSn
Bohr magneton value per Mn atom	4·04(6)	4·04	4·14
Curie point (° C)	233	330	—
internuclear separation (kX)	4·38	4·17	4·36
$\frac{R}{r} = \frac{\text{atomic diam.}}{\text{electron shell diam.}}$ (incomplete shell)	2·98	2·84	2·97

where J = interaction energy per electron pair, θ = Curie temperature, Z = number of nearest neighbours, k = Boltzmann's constant. The Curie temperature is thus a measure of the interaction energy. The relationship between the interaction energy, as measured by the Curie temperature, and the ratio R/r has been quantitatively established from observations on the normal ferromagnetics and their alloys. Slater's (1930) empirical rules indicate a value of about 1·47 kX for the radius of the 3*d* shell in manganese and this value has been used in calculating the values of R/r in table 3. It will be seen that the values of R/r lie between 2·84 and 2·98, and are considerably larger than those for the normal ferromagnetics. According to the Stoner-Slater hypothesis, the interaction energies should be much weaker owing to the greater separation of the atoms, and this should be manifested by lower Curie temperatures, as is in fact the case. Now gadolinium has been found to be ferromagnetic with a Curie temperature of ~ 16° C with an R/r value of 3·1. Clearly this indicates the approach to the upper limit of R/r at which interaction is insufficient to stabilize the ferromagnetic state. It is significant therefore that the increase in the value of R/r from 2·84 in Cu_2MnAl to 2·98 in Cu_2MnIn corresponds with a fall of ~ 97° C in the Curie temperature. These two alloys are strictly analogous as regards valencies, structures and electron concentration, and the figures suggest clearly that the critical value is being approached at which ferromagnetism will disappear.

Objection has been taken (e.g. Stoner 1934) to the view that the magnetic properties of the Heusler alloys are the result of interaction between manganese atoms, on the grounds that the distances are too great. Slater (1930) suggested that in a Heusler alloy the copper atoms might exist in a divalent form with an incomplete *d* shell so that interaction could occur between manganese and copper atoms.*

---

* Interaction can only take place between incomplete shells and so a univalent copper atom with a complete $(3d)^{10}$ sub-group could not interact with manganese to produce ferromagnetism.

This hypothesis appears less probable than that of normal interaction between manganese atoms, because it is difficult to imagine the existence of divalent copper atoms in the presence of highly electro-positive metals like aluminium, whose valency electrons would have a strong tendency to fill the vacancies in the *d* shells of the copper atoms.

If the hypothesis of normal interaction between manganese atoms be accepted, the resulting value of approximately 4 Bohr magnetons per manganese atom would mean that each manganese atom had about six electrons in the 3*d* shell (four unpaired, and two paired), leaving approximately one electron free to join the valency electrons of the crystal lattice. This would imply an almost univalent manganese in agreement with the quite independent evidence obtained from the study of the phase-boundaries (Hume-Rothery 1948). It is also to be noted that Potter (1928) showed that the directional characteristics of the magnetic properties of single crystals of Cu_2MnAl were those of the face-centred cubic arrangement of the manganese atoms† in figure 1, and thus provided further evidence for manganese-manganese interaction. On the whole, therefore, the hypothesis of normal interaction between manganese atoms appears the least improbable assumption since it implies a valency agreeing with that deduced from the phase-boundaries, whilst the difference in Curie points of Cu_2MnAl and Cu_2MnIn suggests that the interatomic distances responsible for the ferromagnetic interaction is almost at its limit.

[*Note added in proof*, 14 December 1948.] Since the proofs of the above paper were corrected, the authors' attention has been drawn to a note by Valentiner (1947). Details of the work are not given, but the lattice spacing 6·2 Å and a Curie temperature of roughly 270° C are in reasonable agreement with the results of the present work.

REFERENCES

Axon, H. J. & Hume-Rothery, W. 1948 *Proc. Roy. Soc.* A, **193**, 1.
Bradley, A. J. & Rodgers, J. W. 1936 *Proc. Roy. Soc.* A, **144**, 340.
Carapella, L. A. & Hultgren, R. 1941 *Tech. Publ. Amer. Inst. Min. Engrs*, no. 1405.
Fallot, M. 1935 Thesis, Strasbourg.
Heusler, O. 1934 *Ann. Phys., Lpz.*, **19**, 155.
Hume-Rothery, W. 1948 *Phil. Mag.* **39**, 89.
Hume-Rothery, W. & Powell, H. M. 1935 *Z. Krist.* **91**, 23.
Hume-Rothery, W., Reynolds, P. W. & Raynor, G. V. 1940 *J. Inst. Met.* **66**, 191.
Potter, H. M. 1927–9 *Proc. Phys. Soc.* **40/41**, 135.
Slater, J. C. 1930 *Phys. Rev.* **36**, 57.
Stoner, E. C. 1930 *Phil. Mag.* **10**, 27.
Stoner, E. C. 1934 *Magnetism and Matter*, p. 525. London: Methuen.
Sucksmith, W. 1939 *Proc. Roy. Soc.* A, **170**, 551.
Valentiner, S. 1947 *Naturwissenschaften*, **34**, 123.
Valentiner, S. & Becker, G. 1933 *Z. Phys.* **83**, 371.
Valentiner, S. & Puzicka, I. 1947 *Metallforschung*, **2**, 127.
Weiss, P. & Forrer, R. 1929 *Ann. Phys., Paris*, **12**, 279.

---

† In figure 1, each set of positions *a*, *b*, *c* or *d* considered alone occupies the face-centred cubic arrangement.

Non-central interactions between neutron and proton

By Tsi-Ming Hu and H. S. W. Massey, F.R.S., *University College, London*

(*Received* 18 *May* 1948—*Revised* 30 *August* 1948)

Calculations are carried out concerning the magnetic moment of the deuteron, the scattering of neutrons with energies up to 20 MeV by protons and the photodisintegration of deuterons by high energy (17 to 20 MeV) γ-rays assuming various non-central interactions between neutron and proton of the form

$$\mathscr{V}=\begin{matrix}-\tfrac{1}{3}\boldsymbol{\tau}_1\cdot\boldsymbol{\tau}_2\\ -\tfrac{1}{2}(1+\boldsymbol{\tau}_1\cdot\boldsymbol{\tau}_2)\\ 1\end{matrix}\left\{-\frac{\hbar^2}{M}\frac{a}{r_2^0}\left[1+\tfrac{1}{2}g(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2-1)+\gamma\left(\frac{3\boldsymbol{\sigma}_1\cdot\mathbf{r}\,\boldsymbol{\sigma}_2\cdot\mathbf{r}}{r^2}-\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2\right)\right]\right\}\,V(r/r_0),$$

where $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ are the spin and isotopic spin operators of the particles, r the distance between them. For three 'shape' functions V (spherical well, exponential and Yukawa forms) the constants a, g, γ are determined for different values of the range r_0, so as to give correctly the binding energy of the ground state, and the quadrupole moment, of the deuteron and the low velocity neutron-proton cross-section. Calculations are then carried out for chosen values of r_0 and the three forms of the exchange operator for each shape. A comparison is made of the ranges derived from different data for each of the shapes considered.

INTRODUCTION

Although it seems very probable that the meson must play some essential role in the eventual theory of nuclear forces it is still far from clear what form this will take. Meanwhile it is possible to proceed in two distinct ways. One may either attempt to seek theoretically an improved field theory from which to derive the nuclear forces or may proceed empirically to obtain information about the forces from experimental investigations. For the latter purpose it is necessary to carry out calculations of the results to be expected for a wide variety of assumptions. A considerable programme of work to this end has already been carried out, assuming the force between neutron and proton to be central in character (Massey & Buckingham 1937; Kittel & Breit 1939; Hulthén 1943, 1944; Pais 1946; Fröhlich, Ramsey & Sneddon 1947 and Ramsey 1947). However, the discovery of the quadrupole moment of the deuteron (Kellogg, Rabi, Zacharias & Ramsey 1939) directed attention to the existence of a non-central interaction between a neutron and a proton. A term of this character can be obtained from vector or pseudoscalar meson field theory but proves to be highly singular at the origin. To remove this singularity, Møller & Rosenfeld (1940) proposed a combined vector and pseudoscalar meson theory which gave a purely central interaction in the first approximation. The quadrupole terms are then supposed to arise in higher approximations. This procedure is formally convenient but awaits detailed confirmation. Moreover, the question whether the singularity does or does not reappear in the higher approximations still seems to be undecided. On the other hand, Bethe (1940) has attempted to develop a theory in which the purely central force term does not appear and the singularity in the remaining non-central force is removed by an empirical cut-off method. Apart from the doubtful significance of the cut-off method other difficulties arise which reduce the applicability of the theory (Volkhoff 1942). An alternative procedure has been introduced by Rarita & Schwinger (1941) which does not attempt to follow the predictions of meson theory in any detail but merely starts from an interaction in the form of a linear combination of terms, involving the spin and isotopic spin vectors, which have the correct transformation characteristics. These terms represent the variation of the interaction with the total spin and with the angles between the spin and relative position vectors. They are then associated with an empirically chosen function of the distance between the particles which involves a range and intensity parameter. The constants in the linear combination, together with those specifying the range and intensity of the forces, are then to be determined from the experimental results. These include the binding energy, electric quadrupole moment and magnetic dipole moment of the deuteron, the collisions of neutrons and of protons with protons, the disintegration of the deuteron by γ-rays and the binding energies of other light nuclei such as H^3, He^3 and He^4. As the adequacy of this data grows it becomes possible to do more than determine the constants. Different sets of data may be used for this purpose and the consistency of the results obtained in these various ways may be checked. In the absence of a satisfactory check the assumed type of interaction would then have to be discarded. So far, detailed calculations on these lines have only been carried out for a limited range of experimental conditions, for forces with a radial dependence of spherical well type with a range of $2{\cdot}8 \times 10^{-13}$ cm. In this paper we describe an extension of this work to deal with a wider range of experimental conditions and other ranges and types of radial variation of the force.

1. ASSUMED FORM OF FUNDAMENTAL INTERACTION

We have adopted the same general forms as those assumed by Rarita & Schwinger (1941). If $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$ are the nuclear spin vector operators, $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ the corresponding isotopic spin operators and $\mathbf{r}$ is the relative position vector of the particles, then, in order to possess the correct transformation properties, the neutron-proton interaction must be a linear combination of the operators

$$1,\ \boldsymbol{\sigma}_1 . \boldsymbol{\sigma}_2,\ \boldsymbol{\tau}_1 . \boldsymbol{\tau}_2,\ S_{12} = \frac{3\boldsymbol{\sigma}_1 . \mathbf{r}\, \boldsymbol{\sigma}_2 . \mathbf{r}}{r^2} - \boldsymbol{\sigma}_1 . \boldsymbol{\sigma}_2.$$

We assume for the interaction the expression

$$\mathscr{V} = -\frac{\hbar^2}{M}\frac{a}{r_0^2} P\left\{1 + \tfrac{1}{2}g(\boldsymbol{\sigma}_1 . \boldsymbol{\sigma}_2 - 1) + \gamma\left(\frac{3\boldsymbol{\sigma}_1 . \mathbf{r}\, \boldsymbol{\sigma}_2 . \mathbf{r}}{r^2} - \boldsymbol{\sigma}_1 . \boldsymbol{\sigma}_2\right)\right\} V(r/r_0), \tag{1}$$

where M is the mass of the nucleon and r_0 is the range of the force. $V(r/r_0)$ defines the radial dependence and will be assumed to be either the 'spherical well'

$$\begin{aligned} V(x) &= 1 \quad (x < 1), \\ &= 0 \quad (x > 1), \end{aligned}$$

the exponential function e^{-2x} or the Yukawa potential e^{-x}/x. P determines the exchange character and we shall consider the cases

$$P_{\mathrm{I}} = -\tfrac{1}{3}\boldsymbol{\tau}_1 \cdot \boldsymbol{\tau}_2, \tag{2}$$

$$P_{\mathrm{II}} = -\tfrac{1}{2}(1+\boldsymbol{\tau}_1 \cdot \boldsymbol{\tau}_2), \tag{3}$$

$$P_{\mathrm{III}} = 1. \tag{4}$$

These forms correspond to symmetrical, charged and neutral meson theory respectively. We do not, however, use the radial dependence of the force which is given by these theories.

The constant a measures the strength of the interaction, g its spin-dependent part and γ the non-central force. For simplicity g and γ are assumed constant.

2. EXPERIMENTAL DATA USED TO DETERMINE THE INTERACTION CONSTANTS

For an assumed shape, given by $V(r/r_0)$, there are four constants a, g, γ and r_0. Three relations between these constants may be obtained without ambiguity from the binding energy of the deuteron (2·19 MeV), the cross-section (20×10^{-24} cm.^{2}) for collision between thermal neutrons and protons (essentially determined by the energy of the ^{1}S state of the deuteron) and the quadrupole moment of the deuteron ($2{\cdot}73 \times 10^{-27}$ cm.^{2}). These relations are independent of the nature of the exchange operator so they apply to all three forms.

The fourth relation can be obtained in principle from the cross-section for the scattering of medium energy (1 to 6 MeV say) neutrons by protons, from the cross-section for disintegration of the deuteron by γ-rays, from the cross-sections for collision of thermal or lower energy neutrons with parahydrogen molecules, and from the magnetic moment of the deuteron. The difficulty in the first three cases is one of obtaining sufficiently precise experimental data. On the other hand, although the magnetic moments of the proton, neutron and deuteron are now known with great precision, there is some uncertainty in the theory of the deuteron moment which makes the fourth relation derived from the magnetic moment data somewhat unreliable.

In view of this position one of the four constants, the range r_0, has been treated as a variable parameter and calculations carried out for more than one range for each shape V. To assist in the choice of the values of r_0 for detailed calculation, information, of a less direct character than that discussed above, may be used (see § 7).

3. DETERMINATION OF THE CONSTANTS a, g AND γ FOR A GIVEN RANGE

3·1. *Energy levels in the non-central field*

The energy levels of the two body system with an interaction energy of the forms (1), (2) have been discussed by Bethe (1940) and by Rarita & Schwinger (1941). The Hamiltonian commutes with the total angular momentum and with the magnitude of the total spin and is invariant against inversion. On the other hand, the orbital angular momentum and the components of the total spin are not constants of the motion. The four quantum numbers specifying a state may therefore be taken as J, determining the total angular momentum, M its z-component, S the total spin and the parity.

As S is a quantum number the states divide into a singlet and a triplet system. For the former, as $S = 0$, $J = l$ the orbital angular momentum quantum number, which is thus a true quantum number for the singlet states. This corresponds to the fact that the non-central interactions, averaged over the spin wave function $\alpha(1)\beta(2)-\alpha(2)\beta(1)$ for a singlet state, vanishes. The effective potential $^1\mathrm{V}$ which determines the singlet states is therefore spherically symmetrical but, when the isotopic spin operators are present, it depends on the parity. With the forms (1), (2), (3) and (4) we have the following expressions for $^1\mathrm{V}$. For even states and all three interactions

$$^1\mathrm{V}_{\mathrm{even}} = -\frac{\hbar^2}{M}\frac{a}{r_0^2}(1-2g)\,V(r/r_0), \tag{5}$$

while for odd states we have, for interaction I $^1\mathrm{V}_{\mathrm{odd}} = -3\,{}^1\mathrm{V}_{\mathrm{even}}$, interaction II $^1\mathrm{V}_{\mathrm{odd}} = -{}^1\mathrm{V}_{\mathrm{even}}$ and interaction III $^1\mathrm{V}_{\mathrm{odd}} = {}^1\mathrm{V}_{\mathrm{even}}$.

For the triplet states the problem is somewhat more complicated. We first note the effect of the isotopic spin operators on the effective interaction for states of different parity. We have now for the respective interactions I, II and III

$$^3\mathrm{V}_{\mathrm{odd}} = -\tfrac{1}{3}\,{}^3\mathrm{V}_{\mathrm{even}}, \quad {}^3\mathrm{V}_{\mathrm{odd}} = -{}^3\mathrm{V}_{\mathrm{even}}, \quad {}^3\mathrm{V}_{\mathrm{odd}} = {}^3\mathrm{V}_{\mathrm{even}}. \tag{6}$$

For a particular triplet state, characterized by J, M, there are three possible l values, $l = J$, $J \pm 1$, and three possible values of the magnetic quantum number m, $m = M$, $M \pm 1$. The wave function ψ_{JM} may therefore be written as the linear combination

$$\psi_{JM} = \sum_{l=J-1}^{J+1} r^{-1} f_{JlM} F_{JlM}(\theta,\phi,s_1,s_2), \tag{7}$$

where
$$F_{JM} = \sum_{m=M-1}^{M+1} c_{JlMm} Y_{lm}(\theta,\phi)\,\chi_{M-m}(s_1,s_2). \tag{8}$$

Here $Y_{lm} = \left\{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}\right\}^{\frac{1}{2}} P_l^m(\cos\theta)\,\mathrm{e}^{im\phi}$, a normalized tesseral harmonic and χ_{M-m} is the spin wave function which takes the usual forms:

$$\chi_1 = \alpha(1)\alpha(2), \quad \chi_{-1} = \beta(1)\beta(2), \quad \chi_0 = 2^{-\frac{1}{2}}\{\alpha(1)\beta(2)+\alpha(2)\beta(1)\}. \tag{9}$$

r, θ, ϕ are the relative polar co-ordinates of the two particles and s_1, s_2 their spin co-ordinates.

The ratios of the constants c_{JlMm} are determined from the condition that ψ_{JM} should be a proper function for the total angular momentum operator J^2, and they may be completely determined (apart from an arbitrary sign) from the normalization condition

$$\sum_{m=M-1}^{M+1} c^2_{JlMm} = 1.$$

TABLE 1. VALUES OF THE COEFFICIENTS c_{JlMm}

	$m = M-1$	$m = M$	$m = M+1$
$l = J-1$	$\left\{\frac{(J+M)(J+M-1)}{2J(2J-1)}\right\}^{\frac{1}{2}}$	$\left\{\frac{J^2-M^2}{J(2J-1)}\right\}^{\frac{1}{2}}$	$\left\{\frac{(J-M)(J-M-1)}{2J(2J-1)}\right\}^{\frac{1}{2}}$
$l = J$	$-\left\{\frac{(J+M)(J-M+1)}{2J(J+1)}\right\}^{\frac{1}{2}}$	$\frac{M}{\{J(J+1)\}^{\frac{1}{2}}}$	$\left\{\frac{(J-M)(J+M+1)}{2J(J+1)}\right\}^{\frac{1}{2}}$
$l = J+1$	$\left\{\frac{(J-M+2)(J-M+1)}{2(J+1)(2J+3)}\right\}^{\frac{1}{2}}$	$-\left\{\frac{(J+1)^2-M^2}{(J+1)(2J+3)}\right\}^{\frac{1}{2}}$	$\left\{\frac{(J+M+2)(J+M+1)}{2(J+1)(2J+3)}\right\}$

As there is no coupling between states of different parity the state with $l = J$ is not coupled to those with $l = J \pm 1$ and, for interaction III, the radial function f_{JJM} satisfies

$$\left[\frac{d^2}{dr^2}-\frac{l(l+1)}{r^2}-\kappa^2+ar_0^{-2}(1+2\gamma)\,V(r/r_0)\right]f_{JJM} = 0. \tag{10}$$

The same equation holds for the other two interactions if J is even, but if J is odd $a(1+2\gamma)$ must be multiplied by $-\frac{1}{3}$ for interaction I and by -1 for interaction II.

For the cases $l = J \pm 1$ and interaction III, we have two simultaneous equations of the form

$$\left[\frac{d^2}{dr^2}-\frac{l(l+1)}{r^2}-\kappa^2\right]f_{JlM}+ar_0^{-2}\,V(r/r_0)\{[1+\gamma t_{Jl}]\,f_{JlM}+\gamma t_{Jll'}f_{Jl'M}\} = 0$$
$$(l, l' = J \pm 1), \tag{11}$$

where the coefficients t_{Jl}, $t_{Jll'}$ are given by (Bethe 1940):

$$t_{J=l+1,l} = -\frac{2(J-1)}{2J+1}, \quad t_{J=l-1,l} = -\frac{2(J+2)}{2J+1}, \quad t_{Jll'} = \frac{6J^{\frac{1}{2}}(J+1)^{\frac{1}{2}}}{2J+1}. \tag{12}$$

The same equations hold for the interactions I and II if J is odd but if J is even $aV(r/r_0)$ must be multiplied by $-\frac{1}{3}$ for interaction I and by -1 for interaction II. It will be noted that in the special case $J = 0$ the equations reduce again to a single equation, i.e. for the $^3\mathrm{P}_0$ state.

3·2. *Determination of the constants a, g and γ*

The ground state of the deuteron is the triplet state with $J = 1$, being a combination of $^3\mathrm{S}$ and $^3\mathrm{D}$ terms. The differential equations for the radial functions are therefore as given by (11) with $J = 1$; $l = 0, 2$, the same for all these interactions, viz.

$$\left.\begin{aligned}\left[\frac{d^2}{dx^2}-\kappa'^2+aV(x)\right]f_{10} &= -2\sqrt{2}\,\gamma aV(x)\,f_{12},\\ \left[\frac{d^2}{dx^2}-\kappa'^2-\frac{6}{x^2}+(1-2\gamma)\,aV(x)\right]f_{12} &= -2\sqrt{2}\,\gamma aV(x)\,f_{10},\end{aligned}\right\} \tag{13}$$

where $x = r/r_0$ and $\kappa'^2 = \kappa^2 r_0^2$.

If the functions are normalized so that

$$\int_0^\infty (f_0^2+f_2^2)\,dx = 1, \tag{14}$$

the quadrupole moment Q of the deuteron is given by

$$\frac{r_0^2}{20}\int_0^\infty (2\sqrt{2}\, f_0 f_2 - f_2^2)\, x^2 dx = Q. \tag{15}$$

Given the form of V and having chosen a value for the range r_0, the problem is to determine a and γ so that, if f_0 and f_2 are to be proper functions, $\kappa^2 = 5{\cdot}24 \times 10^{24}$ cm.$^{-2}$ and $Q = 2{\cdot}73 \times 10^{-27}$ cm.^{2}. An iterative method of approximation, described in appendix I, was found suitable for this purpose.

Figure 1 illustrates the variation of a, γ and $a\gamma$ with r_0 for the spherical well potential. The full line curves have been obtained by actual numerical calculation but the dotted extrapolations are only intended to indicate general features. It will be seen that for $r_0 < r_1$, no real values of a and γ will give a large enough quadrupole moment. For $r_1 < r_0 < r_2$ the central force is attractive, the non-central repulsive. As r_0 approaches r_2 the depth constant a tends to infinity. In the next region, $r_2 < r_0 < r_3$, the signs of the central and non-central forces are reversed. The central force decreases to zero as r_0 approaches r_3. Thus for this range the force is entirely non-central. Finally, for $r_0 > r_3$, both components become attractive and the importance of the non-central force decreases with increasing r_0.

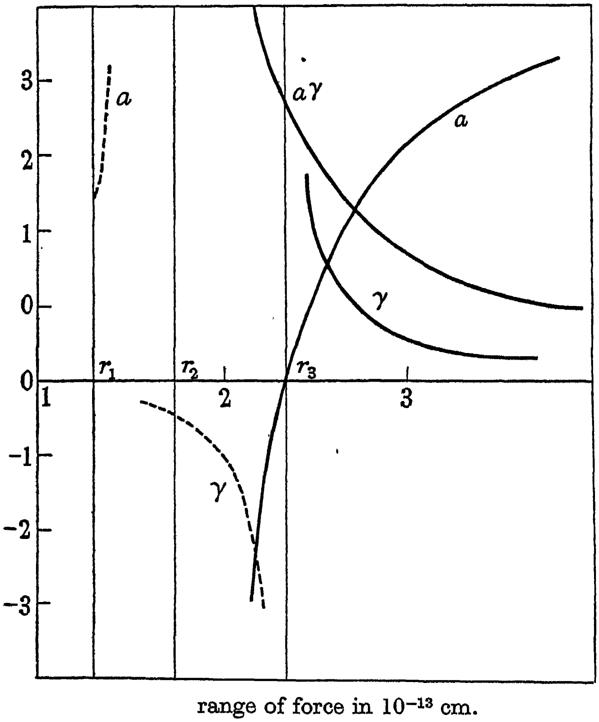


FIGURE 1. Variation of the constants a, γ and $a\gamma$ with range r_0 for an interaction of spherical well shape in order to give the correct binding energy and quadrupole moment of the deuteron.

This general behaviour is independent of the shape of the function $V(r/r_0)$. Numerical values of a and γ for the Gaussian, exponential and Yukawa potentials are given in table 2, together with results for the spherical well.

The constant g is determined by the energy of the ${}^1S$ (virtual) state of the deuteron which is not affected by the non-central part of the potential. Taking the energy of the ${}^1S$ level as 75,000 eV (derived from the measured collision cross-section 20×10^{-24} cm.^{2} of protons for thermal neutrons) the values of g given in table 2 are obtained for the various cases.

TABLE 2. INTERACTION CONSTANTS

shape of interaction	range of force, r_0 (10^{-13} cm.)	κr_0	a	γ	g	q
spherical well	2·18	0·5	−1·83	−2·61	1·13	0·085
spherical well	2·62	0·6	1·93	1·29	−0·086	0·049
spherical well	3·05	0·7	3·24	0·50	0·156	0·030
spherical well	3·49	0·8	3·97	0·29	0·224	0·019
Yukawa	1·74	0·4	1·94	0·73	0·107	0·026
Yukawa	2·18	0·5	2·43	0·44	0·194	0·017
Yukawa	3·05	0·7	3·09	0·24	0·271	0·010
exponential	1·74	0·4	6·2	0·83	0·074	0·032
exponential	2·18	0·5	8·2	0·44	0·185	0·020
Gaussian	2·18	0·5	3·39	0·59	—	0·029

It was assumed that the ground state of the deuteron is a ${}^3S + {}^3D$-state but it is important to notice that this becomes inconsistent with the values of the constants if the range is too small. Thus, for the interaction I, the ${}^3P_0$ state becomes stable if $r_0 < 2{\cdot}05 \times 10^{-13}$ cm.* and would be lower than ${}^3S + {}^3D$ for $r_0 < 2{\cdot}02 \times 10^{-13}$ cm.* For interaction II the corresponding values are $2{\cdot}52 \times 10^{-13}$ cm.* and $2{\cdot}48 \times 10^{-13}$ cm.* Interaction III similarly gives rise to a stable ${}^3P_1$ state lying lower than ${}^3S + {}^3D$ if the range is less than about $1{\cdot}9 \times 10^{-13}$ cm. This imposes an important restriction on the magnitude of the range which can be adopted with the interactions we have assumed.

4. MAGNETIC MOMENT OF THE DEUTERON

The wave function for the ground state of the deuteron is not entirely of ${}^3S_1$ type but includes an admixture of ${}^3D_1$ component, which contributes an orbital magnetic moment. This contribution is given by

$$-\tfrac{3}{2}(\mu_n + \mu_p - \tfrac{1}{2})\, q, \tag{16}$$

where $q = \int_0^\infty f_{12}^2\, dx$ and μ_n, μ_p are the respective magnetic moments of a free neutron and proton. q is a measure of the proportional admixture of the ${}^3D$ component. It is given as a function of r_0 for the different shapes of potential in table 2. In general q increases as the range decreases because it depends on the charge distribution at rather smaller distances than does the quadrupole moment. As a result to maintain this moment constant as the range decreases, q must be increased.

---

* These values were obtained by interpolation. The absolute values are not reliable to three figures, but the relative positions are correct.

If it can be assumed that the intrinsic magnetic moments of neutron and proton are unaffected by their binding the total magnetic moment of the deuteron would be

$$\mu_D = \mu_n + \mu_p - \tfrac{3}{2}(\mu_n + \mu_p - \tfrac{1}{2})\,q. \tag{17}$$

μ_D, μ_n and μ_p have recently been measured with great precision, the values obtained being, in nuclear magnetons,

$\mu_p = 2{\cdot}7896 \pm 0{\cdot}0008$ (Millman & Kusch 1941),
$\mu_n = -1{\cdot}9108 \pm 0{\cdot}0016$ (Arnold & Roberts 1947),
$\mu_D = 0{\cdot}8565 \pm 0{\cdot}0004$ (Millman & Kusch 1941).

In order that these values should be consistent with (17) q must be taken as $0{\cdot}04 \pm 0{\cdot}003$. This gives for the respective ranges, in units 10^{-13} cm., of the spherical well, exponential and Yukawa interactions 2·8, 1·6 ± 0·05 and 1·4 ± 0·05 respectively.

The difficulty about this determination of range is that (17) ignores the effect of the binding and of relativity on μ_n and μ_p. The first of these would be expected to arise from the distortion of the virtual meson charge round the nucleons by their mutual interaction. Reasons have been given (Villars 1947) why this effect might be negligible in the deuteron, but the position is still not quite definite. Even less certain (Breit & Bloch 1947; Sachs 1947; Primakoff 1947) is the importance of the relativistic correction due to the motion of the nucleons but it may be comparable in importance to that arising from the non-central force. The ranges derived above can therefore be regarded as unreliable, at least until the theory has been fully developed.

5. THE ELASTIC SCATTERING OF NEUTRONS BY PROTONS

A detailed study of the velocity variation of the total cross-section and angular distribution for the scattering of neutrons by free protons at rest is likely to provide information about the exchange character as well as the range of the fundamental interaction. It is also of interest to examine the importance of the non-central terms in determining the character of the scattering cross-sections.

The modifications to the usual scattering formulae due to the non-central force have been discussed by Rarita & Schwinger (1941). When the neutron spin is anti-parallel to that of the proton the only new feature introduced is the dependence of the scattering potential on the parity. The differential cross-section ${}^1I(\theta)\,d\omega$, for collisions in which the kinetic energy E of relative motion is $\hbar^2k^2/M$, is given by

$${}^1I(\theta) = \frac{1}{4k^2}\left|\sum_0^\infty (\mathrm{e}^{2i\eta_l}-1)(2l+1)P_l(\cos\theta)\right|^2,$$

where η_{2s} is such that the asymptotic form of the proper solution of

$$\frac{d^2f_{2s}}{dr^2} + \left\{k^2 + \frac{a(1-2g)}{r_0^2}V(r/r_0) - \frac{2s(2s+1)}{r^2}\right\}f_{2s} = 0$$

is

$$f_{2s} \sim \sin(kr - s\pi + \eta_{2s}).$$

η_{2s+1} is defined in the same way for forces of type III (neutral theory), but for forces of types I and II V is replaced by $-3V$ and $-V$ respectively. The total cross-section ${}^1Q$ is given by

$$ {}^1Q = \frac{4\pi}{k^2}\Sigma(2l+1)\sin^2\eta_l. \tag{18}$$

For the triplet scattering the situation is more complicated. The method of Rarita & Schwinger was developed especially for application to the scattering of waves with low relative angular momentum. We shall derive the general formulae by a method which is exactly parallel to that employed in the theory of scattering by central forces.

We calculate first the cross-section for scattering when the component of total spin, resolved along the direction of incidence, is $M\hbar$. The wave function describing the scattering can then be expanded in the form

$$\Psi = \frac{(4\pi)^{\frac{1}{2}}}{kr}\sum_{J=0}^{\infty}\sum_{l=J-1}^{J+1} f_{JlM}(r)\,F_{Jlm}(\theta,\phi,s),$$

where F_{JlM} is as defined in (4). The functions f_{JlM} will have the asymptotic form

$$f_{JlM} \sim \kappa_{JlM}\sin\left(kr - \tfrac{1}{2}l\pi + \eta_{JlM}\right). \tag{19}$$

The incident wave $e^{ikz}\chi_M(s)$ can be expanded in the form

$$e^{ikz}\chi_M = \frac{(4\pi)^{\frac{1}{2}}}{kr}\sum_{J=0}^{\infty}\sum_{l=J-1}^{J+1} i^l c_{JlM0}(2l+1)^{\frac{1}{2}}\,g_l(r)\,F_{JlM}, \tag{20}$$

where
$$\begin{aligned} g_l &= \left(\tfrac{1}{2}\pi kr\right)^{\frac{1}{2}} J_{l+\frac{1}{2}}(kr) \\ &\sim \sin\left(kr - \tfrac{1}{2}l\pi\right). \end{aligned}$$

We may write then

$$\Psi = e^{ikz}\chi_M + \frac{(4\pi)^{\frac{1}{2}}}{kr}\sum_{J=0}^{\infty}\sum_{l=J-1}^{J+1}\left(f_{JlM} - i^l(2l+1)^{\frac{1}{2}} c_{JlM0}\,g_l\right)F_{JlM}. \tag{21}$$

If Ψ is to describe the scattering correctly the sum must represent asymptotically an outgoing spherical wave. For this we must have

$$\kappa_{JlM} = i^l(2l+1)^{\frac{1}{2}}\exp\left(i\eta_{JlM}\right)c_{JlM0}. \tag{22}$$

The differential cross-section ${}^3I_M(\theta)$ for scattering through an angle θ into the solid angle $d\omega$ is now given by

$$ {}^3I_M(\theta) = \frac{\pi}{k^2}\sum_s\left|\sum_{l=0}^{\infty}\sum_{J=l-1}^{l+1}(2l+1)^{\frac{1}{2}}c_{JlM0}\{\exp\left(2i\eta_{JlM}\right)-1\}F_{JlM}\right|^2 d\Omega, \tag{23}$$

where Σ_s denotes summation over the spin co-ordinates.

On substitution for the c_{JlM0} and F_{JlM} we obtain finally

$$ {}^3I_0(\theta) = {}^3I_{00}(\theta) + {}^3I_{01}(\theta), \quad {}^3I_{\pm1}(\theta) = {}^3I_{12}(\theta) + {}^3I_{11}(\theta) + {}^3I_{10}(\theta),$$

where

$$\left.\begin{aligned}
{}^3I_{00}(\theta) &= \frac{1}{4k^2}\left|\sum_{l=0}^{\infty}\{l\epsilon_{l-1,l,0}+(l+1)\,\epsilon_{l+1,l,0}\}\,P_l(\cos\theta)\right|^2,\\
{}^3I_{01}(\theta) &= \frac{1}{4k^2}\left|\sum_{l=1}^{\infty}\{\epsilon_{l-1,l,0}-\epsilon_{l+1,l,0}\}\,P_l^1(\cos\theta)\right|^2,\\
{}^3I_{11}(\theta) &= \frac{1}{8k^2}\left|\sum_{l=1}^{\infty}\left\{\frac{l-1}{l}\epsilon_{l-1,l,1}+\frac{2l+1}{l(l+1)}\epsilon_{l,l,1}-\frac{l+2}{l+1}\epsilon_{l+1\,l,1}\right\}P_l^1(\cos\theta)\right|^2,\\
{}^3I_{12}(\theta) &= \frac{1}{16k^2}\left|\sum_{l=2}^{\infty}\left\{\frac{1}{l}\epsilon_{l-1,l,1}-\frac{2l+1}{l(l+1)}\epsilon_{l,l,1}+\frac{1}{l+1}\epsilon_{l+1,l,1}\right\}P_l^2(\cos\theta)\right|^2,\\
{}^3I_{10}(\theta) &= \frac{1}{16k^2}\left|\sum_{l=0}^{\infty}\{(l-1)\,\epsilon_{l-1,l,1}+(2l+1)\,\epsilon_{l,l,1}+(l+2)\,\epsilon_{l+1,l,1}\}\,P_l(\cos\theta)\right|^2,
\end{aligned}\right\} \quad (25)$$

and

$$\epsilon_{JlM} = \exp(2i\eta_{JlM})-1.$$

The total cross-section Q is given by

$$Q = \tfrac{1}{4}[{}^1Q+{}^3Q_{M=0}+{}^3Q_{M=+1}+{}^3Q_{M=-1}],$$

where

${}^1Q$ as given in (18),

$${}^3Q_{M=0} = \frac{\pi}{k^2}\sum_{l=0}^{\infty}\{l\,|\,\epsilon_{l-1,l,0}\,|^2+(l+1)\,|\,\epsilon_{l+1,l,0}\,|^2\},$$

$${}^3Q_{M=\pm1} = \frac{\pi}{2k^2}\sum_{l=0}^{\infty}\{(l-1)\,|\,\epsilon_{l-1,l,1}\,|^2+(2l+1)\,|\,\epsilon_{l,l,1}\,|^2+(l+2)\,|\,\epsilon_{l+1,l,1})^2\}.$$

When the coupling terms vanish, as is the case when $J = l$ and $J = 0$, the phases η may be derived at once from the asymptotic form of the proper solution of the differential equation for f_{JJM} and f_{01M}. If, however, $J = l \pm 1$ the functions $f_{J,J+1,M}$, $f_{J,J-1,M}$ are derived as proper solutions of two simultaneous second order equations. Each function will just be obtained as a linear combination of two functions involving two undetermined constants A_{JM}, B_{JM} independent of l. These constants and the two phases $\eta_{J,J+1,M}$, $\eta_{J,J-1,M}$ may be determined by the condition that f_{JlM} have the asymptotic form (19) with κ_{JlM} as given in (22). In doing so a dependence of the phase η on M will be found, even though M does not appear in the appropriate coupled equations. It is also noteworthy that the phases derived from the coupled equations are in general complex corresponding to fluctuation of amplitude between the two states.

The cross-sections (24) and (26) must finally be averaged over the possible M values to give ${}^3I(\theta)$ and ${}^3Q$ the cross-sections for scattering when the neutron and proton spins are parallel. Finally, for comparison with observation the cross-sections must be averaged also over the neutron spin to give

$$I(\theta) = \tfrac{1}{4}({}^1I+{}^3I_0+{}^3I_1+{}^3I_{-1}), \quad Q = \tfrac{1}{4}{}^1Q+\tfrac{3}{4}{}^3Q. \quad (26)$$

At a given neutron energy the phases η_{JlM} will be small for such values of l that

$$kr_0 \sim l,$$

where r_0 is the range of the force. For sufficiently low energies the only important phases will be those for which $l = 0$. At any particular energy the convergence of the series of partial cross-sections is best for the spherical well potential owing to the absence of any interaction at long range in that case.

Over a large part of the neutron energy range covered by the present calculations the major contributions come from the zero and first order phases but even then a great number (about 500) of phases have to be determined to cover the variety of assumed interactions. The methods actually used in determining these are described in appendix II.

The zero order phases do not depend on the nature of the exchange operator and, being determined largely by the binding energy of the respective ^{3}S and ^{1}S states of the deuteron, are fairly insensitive to the shape of the potential and to the range for a particular shape, especially at the lower neutron energies. On the other hand, the first order phases do depend quite strongly on all these features.

For neutrons with energy of this order and a central interaction, the P phases given by exchange forces of either type I or type II are of opposite sign to those given by ordinary forces of type III which are always positive. With non-central forces of the form assumed this is no longer the case because the $^3P_0$ phase for each type of interaction has the opposite sign to the remaining P phases. As it is associated with a small statistical weight it does not completely vitiate the discrimination between the predicted scattering with exchange and ordinary forces. Its influence is most marked for type II forces with a large non-central component, i.e. a short range. For the exponential and Yukawa forms of $V(r/r_0)$ it is necessary to allow for the contributions from D and even F phases.

Using the calculated phases the total cross-sections Q and angular distributions $I(\theta)$ were calculated for neutron energies of 2·13, 5·20, 13·3 and 20·8 MeV for a variety of interactions.

Table 3 gives the calculated cross-sections Q. For comparison with central force theories values obtained by assuming a purely central interaction of the same range and shape, with depth adjusted to give the correct binding of the appropriate deuteron state (^{3}S or ^{1}S), are also given together with experimental values (Salant & Ramsey 1940; Amaldi *et al.* 1943; Sherr 1945; Bailey *et al.* 1946; Sleater 1947) whenever available. It appears from a study of this table that the non-central character of the force has no marked effect on the total cross-section except for ordinary forces (at the higher energies) and exchange forces of type II at the smallest range.

Comparison with the experimental results shows clearly that non-central interactions of type III predict too large cross-sections for higher neutron energies. Owing to the lack of sensitivity of the cross-section to change of range of the force for this type of interaction,* a large reduction in range is necessary to reduce the calculated cross-section to the experimental value, even in the cases of the exponential well and Yukawa potentials where the discrepancy is not so marked as in the case of the spherical well. Apart from other reasons against this reduction of range in the case of non-central interactions, a much smaller range will predict,

---

* The decreases in the phases due to the reduction in range are, to some extent, balanced by the necessary increase in the non-central part of the interaction.

TABLE 3. CALCULATED SCATTERING CROSS-SECTIONS

neutron energy (MeV)	shape of interaction	range of force (10^{-13} cm.)	total cross-sections Q (10^{-24} cm.^{2}) non-central type I	non-central type II	non-central type III	central type I	central type II	central type III	experimental value (10^{-24} cm.^{2})	asymmetry factor $I(\pi)/I(\frac{1}{2}\pi)$ non-central type I	non-central type II	non-central type III	central type I	central type II	central type III
2·1	spherical well	2·6	—	3·05	—	—	2·98	—	2·95–2·80*	1·00	1·03	1·05	1·00	0·99	1·03
	spherical well	3·0	—	3·21	—	—	3·08	—		1·00	0·99	1·06	1·00	0·98	1·06
	spherical well	3·5	—	3·20	—	—	3·21	—		1·00	0·97	1·11	0·99	0·97	1·10
	exponential	1·74	—	2·86	—	—	2·89	—		—	—	—	—	—	1·06
	Yukawa	1·74	—	—	—	—	2·78	—		—	—	—	—	—	—
5·2	spherical well	2·6	1·67	1·82	1·70	1·67	1·67	1·67	1·6 –1·5*	1·01	0·93	1·07	1·00	0·98	1·07
	spherical well	3·0	1·80	1·80	1·86	1·77	1·77	1·79		1·01	0·99	1·09	1·00	0·97	1·11
	spherical well	3·5	1·81	1·81	1·99	1·81	1·81	1·83		1·02	0·97	1·20	1·01	1·97	1·11
	exponential	1·74	1·62	1·63	1·66	1·62	1·62	1·64		1·04	1·00	1·13	—	—	1·15
	Yukawa	1·74	—	—	—	1·57	1·57	1·59		—	—	—	(1·04)	—	—
13·3	spherical well	2·6	0·728	0·914	1·020	0·726	0·728	0·791	0·8 –0·7†	1·09	0·57	0·64	1·08	1·06	1·01
	spherical well	3·0	0·737	0·745	1·107	0·732	0·737	0·851		1·14	1·05	0·61	1·15	1·13	0·96
	spherical well	3·5	0·711	0·721	1·216	0·710	0·722	1·056		1·31	1·30	0·62	1·34	1·38	0·96
	exponential	1·74	0·734	0·767	0·903	—	0·735	0·785		1·21	0·96	0·79	—	1·14	1·02
	Yukawa	1·74	0·746	0·779	0·922	(0·74)	—	—		1·45	1·03	0·68	(1·46)	—	—
20·8	spherical well	2·6	0·446	0·558	0·772	0·439	0·443	0·528	0·5 –0·4†	1·23	0·71	0·48	1·27	1·30	0·71
	spherical well	3·0	0·434	0·454	0·783	0·430	0·439	0·645		1·45	1·25	0·50	1·55	1·64	0·82
	spherical well	3·5	0·407	0·424	0·843	0·409	0·430	0·852		2·12	2·26	0·56	2·41	2·75	1·05
	exponential	1·74	—	—	—	—	0·475	0·531		—	—	—	—	1·64	0·83
	Yukawa	1·74	0·504	0·535	0·661	(0·51)	—	—		2·11	1·12	0·43	(2·25)	—	—
				I					II				III		

			angular distribution $2k^2I(\theta)$ ($x = \cos\theta$)		
13·3	spherical well	2·6	$1\cdot83-0\cdot08x+0\cdot07x^2-0\cdot01x^3$	$2\cdot30+1\cdot07x+0\cdot07x^2-0\cdot01x^3$	$2\cdot37+1\cdot49x+0\cdot68x^2+0\cdot04x^3$
	exponential	1·74	$1\cdot81-0\cdot14x+0\cdot17x^2-0\cdot06x^3+0\cdot01x^4$	$1\cdot93+0\cdot19x+0\cdot05x^2-0\cdot04x^3+0\cdot01x^4$	$2\cdot07+0\cdot95x+0\cdot68x^2+0\cdot14x^3+0\cdot01x^4$
	Yukawa	1·74	$1\cdot80-0\cdot10x+0\cdot19x^2-0\cdot14x^3-0\cdot14x^4$ $-0\cdot11x^5+0\cdot10x^6-0\cdot02x^7+0\cdot01x^8$	$1\cdot97+0\cdot13x-0\cdot05x^2+0\cdot15x^3+0\cdot09x^4$ $-0\cdot16x^5+0\cdot09x^6-0\cdot03x^7+0\cdot01x^8$	$2\cdot06+1\cdot05x+0\cdot66x^2+0\cdot35x^3+0\cdot28^4$ $+0\cdot26x^5+0\cdot09x^6+0\cdot03x^7+0\cdot01x^8$
20·8	spherical well	2·6	$1\cdot70-0\cdot12x+0\cdot23x^2-0\cdot03x^3+0\cdot01x^4$	$2\cdot15+0\cdot90x+0\cdot22x^2-0\cdot05x^3+0\cdot01x^4$	$2\cdot64+2\cdot58x+1\cdot29x^2+0\cdot09x^3+0\cdot01x^4$
	Yukawa	1·74	$1\cdot74-0\cdot13x+0\cdot45x^2-0\cdot18x^3+0\cdot26x^4$ $-0\cdot33x^5+0\cdot39x^6-0\cdot11x^7+0\cdot07x^8$	$2\cdot03+0\cdot06x+0\cdot07x^2+0\cdot44x^3+0\cdot11x^4$ $+0\cdot07x^5+0\cdot36x^6-0\cdot20x^7+0\cdot07x^8$	$2\cdot20+2\cdot04x+0\cdot84x^2+0\cdot07x^3+0\cdot49x^4$ $+0\cdot69x^5+0\cdot36x^6+0\cdot20x^7+0\cdot07x^8$

---

* Interpolated from the results of Bailey *et al.* (1946).

† Estimated from the results of Salant & Ramsey (1940); Amaldi *et al.* (1943); Sherr (1945) and Sleater (1947).

in every case, too small cross-sections for low energy neutrons. For interactions of type II, it is also not possible to fit the experimental cross-sections for the neutron energy range 2 to 20 MeV with a spherical well interaction because a range larger than $2{\cdot}6 \times 10^{-13}$ cm. is necessary at the higher energy end and a smaller one to fit the low energy end. For the type I interaction, one can fit the experimental values tolerably well with a spherical well of range nearly $2{\cdot}4 \times 10^{-13}$ cm. With the exponential and Yukawa potentials, the range chosen in the calculation give good agreement with observations for both types of exchange force. More refined measurements would be necessary to discriminate between them.

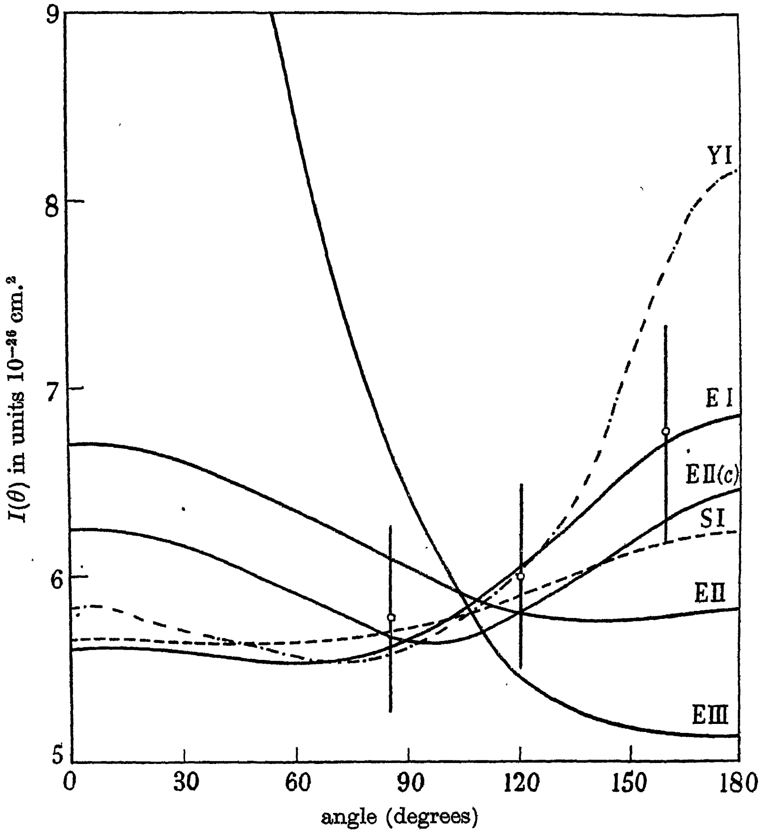


FIGURE 2. Angular distribution $I(\theta)$ per unit solid angle in the C.M. system for scattering of protons by 13·3 MeV neutrons, calculated on different assumptions concerning the interaction. EI exponential form of type I, range $1{\cdot}74 \times 10^{-13}$ cm. EII exponential form of type II, range $1{\cdot}74 \times 10^{-13}$ cm. EII(c) exponential form of type II, range $1{\cdot}74 \times 10^{-13}$ cm., but taken to be wholly central. EIII exponential form of type III, range $1{\cdot}74 \times 10^{-13}$ cm. SI spherical well, of type I, range $2{\cdot}62 \times 10^{-13}$ cm. YI Yukawa form, of type I, range $1{\cdot}74 \times 10^{-13}$ cm. ɸ Experimental observations of Powell & Occhialini (Ramsey 1947) for 13 MeV neutrons with range of error.

The angular distribution of the scattering is more sensitive to the type of force. Table 3 gives calculated values of the ratio $I(\pi)/I(\frac{1}{2}\pi)$ of the scattered intensity at 180° (in the centre-of-mass system) to that at 90° for the same energies and assumed interactions as for Q. It will be seen that measurement of this asymmetry ratio offers much more hope of discrimination between ordinary and exchange type forces.

For this purpose it is clear that neutrons with energies not much less than 13 MeV would have to be used. At this energy and with non-central forces, the ratio for ordinary forces is much smaller than for either exchange types (except type II with the spherical well with range $2{\cdot}6 \times 10^{-13}$ cm. which, in any case, would seem to be excluded from the total cross-section data of table 3). Thus if the asymmetry ratio is found to be nearly equal to unity then 'ordinary' non-central forces would be excluded. It is of interest to note from table 4 that this would not be so if the forces were central. Observations at a higher energy such as 20 MeV would serve to discriminate between non-central forces of exchange type and ordinary central forces.

Detailed evaluation of the expected angular distribution function was also carried out for the various assumed interactions and table 3 gives the coefficients in the expansion of this function in powers of $\cos\theta$ for the two energies of most interest (13·3 and 20·8 MeV). Insufficient experimental data are available at present to enable much choice to be made. In figure 2 the experimental results, with their probable error, obtained by Powell & Occhialini (1947) are compared with various theoretical curves. These observations are definitely against assumption of an ordinary non-central force but unfortunately the position is not yet so definite as this because experimental results have been reported (Amaldi *et al.* 1942) which do not agree with those illustrated in figure 2. On the other hand, measurements by Laughlin & Kruger (1948) with 12 to 13 MeV neutrons do tend to favour the work of Powell & Occhialini.

6. PHOTO-DISINTEGRATION OF THE DEUTERON BY HIGH ENERGY γ-RAYS

The main contribution to the cross-section for disintegration of the deuteron by high energy γ-rays arises from electric dipole transitions. As the initial state is of mixed ^{3}S + ^{3}D type, transitions to the continuous states can take place which are wholly or partially of ^{3}P and ^{3}F character, the latter possibility arising only from the presence of the small ^{3}D admixture in the ground state. As the ^{3}F functions are very small indeed over the range of the ground state wave functions when the energy of relative motion of the disintegrated particle is not too large, we may neglect any such contributions in the present calculations.

Employing the usual perturbation formula for this process, we find for the differential cross-section $I(\theta)$ the formula

$$I(\theta) = \frac{kr_0^2}{12}\frac{e^2}{\hbar c}\frac{E}{(E-E_0)}[\sin^2\theta\{\tfrac{1}{6}\,|\,I_0+I_{20}\,|^2-\tfrac{1}{6}\,|\,I_0\,|^2$$

$$+\tfrac{1}{8}\,|\,3I_1+I_{21}\,|^2-|\,I_1\,|^2+\tfrac{1}{4}\,|\,I_1-I_{21}\,|^2+\tfrac{1}{9}\,|\,I_0-I_{20}\,|^2\}], \quad (27)$$

and for the total cross-section

$$Q = \pi\frac{kr_0^2}{27}\frac{e^2}{\hbar c}\frac{E}{(E-E_0)}\{|\,I_0\,|^2+3\,|\,I_1\,|^2+3\,|\,I_{21}\,|^2+2\,|\,I_{20}\,|^2\}, \quad (28)$$

where $$I_0 = \int_0^\infty rp_0(f_0-\sqrt{2}\,f_2)\,dr, \quad I_1 = \int_0^\infty rp_1\left(f_0+\frac{1}{\sqrt{2}}f_2\right)dr,$$

$$I_{20} = \int_0^\infty rp_{20}\left(f_0-\frac{\sqrt{2}}{10}f_2\right)dr, \quad I_{21} = \int_0^\infty rp_{21}\left(f_0-\frac{\sqrt{2}}{10}f_2\right)dr.$$

f_0/r and f_2/r are the respective ^{3}S and ^{3}D components of the radial wave function for the ground state of the deuteron which is normalized to unity. p_0/r, p_1/r, p_{20}/r p_{21}/r are the respective wave functions for the ^{3}P continuous states

$$({}^3P_0, {}^3P_1, {}^3P_2(M=0) \quad \text{and} \quad {}^3P_2(M=\pm 1)).$$

They are normalized so that $p \sim e^{-i\delta} \sin(kr - \frac{1}{2}\pi + \delta)$ with δ the appropriate phase. δ is related to the corresponding phase η appearing in the theory of elastic scattering by the relation

$$\delta = \eta^*.$$

E is the quantum energy of the incident photon, E_0 the absolute value of the binding energy of the deuteron and $k^2 = M(E-E_0)\, r_0^2/\hbar^2$.

The formulae (27) and (28) reduce to those given by Rarita & Schwinger if the coupling between $^3P_2$ and $^3F_2$ waves is neglected. Since the coupling between $^3P_2$ and $^3F_2$ waves changes appreciably the phase shift for $^3P_2$ waves, it may effect the angular distribution of the disintegrated nucleons. In order to be certain about this point, detailed calculations have been carried out for 17·5 MeV γ-rays, using the Yukawa potential. The results were then compared with those obtained by neglecting their coupling. It was found that the inclusion of their coupling does increase the isotropic component considerably, although the change in total cross-section is much less important (see table 4). Since we are interested in the order of magnitude of this term for different types of interactions rather than its precise evaluation, it is considered sufficient for our purpose to neglect the $^3P_2$-$^3F_2$ coupling in all other cases.

The result of calculations for γ-rays of energy 17·5 MeV and 28·8 MeV are given in table 4. Parallel calculations with central interactions have been carried out and the results included in the same table. It is obvious from the data given in the table that the inclusion of the non-central force has no remarkable effect on the total disintegration cross-sections, except at the higher energies. The chief effect is to introduce an isotropic component in the angular distribution of the ejected particles, which in the central force case varies as $\sin^2\theta$. This effect is only important for type III interaction (ordinary force) unless the range of force is much less than any used in the present calculations. As has been pointed out by Rarita & Schwinger, this difference in angular distribution can be used as an effective means of discriminating between interactions of type III and those of exchange type (I and II). Measurements of total cross-sections will also provide valuable confirmatory evidence in this respect as the ordinary force theory predicts a smaller cross-section than those predicted by forces of type I and II. Discrimination between type I and II interactions would be much more difficult and could not be carried out by photo-disintegration experiments alone, because the predicted angular distribution by these two types of interactions varies considerably with different ranges and shapes of the potential wells, the ranges of these variations nearly overlapping each other. However, precise empirical knowledge in this respect is still very valuable as confirmatory evidence on the nature of nuclear forces. At the time of writing no such experimental evidence is available. In view of the availability of betatron sources, it seems likely that detailed experimental information in this respect will be forthcoming in the near future.

TABLE 4. CALCULATED PHOTO-DISINTEGRATION CROSS-SECTIONS

energy of the γ-rays	shape of	range of force	total cross-section (10^{-28} cm.^{2})						the isotropic component* (A)		
			non-central			central			non-central force		
(MeV)	interaction	(10^{-13} cm.)	type I	type II	type III	type I	type II	type III	type I	type II	type III
17·5	spherical well	2·62	7·23	6·87	4·27	7·53	7·42	5·71	0·019	0·119	0·436
	spherical well	3·05	8·09	7·86	3·99	8·20	8·26	4·00	0·013	0·037	0·341
	spherical well	3·49	8·96	9·00	2·80	9·29	8·77	1·72	0·010	0·017	0·292
	exponential	1·74	7·33	7·38	4·45	—	—	—	0·018	0·071	0·199
	Yukawa	1·74	7·25	7·67	4·59	—	—	—	0·017	0·058	0·106
			(7·26)†	(7·56)	(4·20)				(0·020)	(0·082)	(0·131)
28·8	spherical well	2·62	5·59	5·70	2·70	3·96	3·85	1·14	0·052	0·123	0·587
	spherical well	3·05	5·81	6·29	1·51	4·12	4·72	0·56	0·032	0·064	0·628
	spherical well	3·49	6·02	7·05	0·70	4·60	6·10	0·24	0·023	0·034	0·684

* The angular distribution of the ejected nucleons is given by $A+\sin^2\theta$.

† Values inside the brackets were calculated without neglecting the coupling between $^3P_2$ and $^3F_2$ waves.

7. DISCUSSION. THE RANGE OF THE NEUTRON-PROTON FORCE*

Although the experimental data at present available are insufficient to permit an adequate check of the self-consistency of any assumed neutron-proton interaction there is already enough to make worthwhile an examination of the position in this regard. One such check is to compare, for a given shape of interaction, the ranges required to give agreement with independent observed properties, choosing as far as possible those properties which are insensitive to the exchange character of the force so that the check is independent of this aspect. For this purpose we give separately in table 5 for each of the spherical well, exponential, Yukawa and Gaussian interactions a number of ranges deduced by use of the following properties:

(*a*) Magnetic moment of the deuteron.

(*b*) Scattering of 1 to 6 MeV neutrons by protons.

(*c*) Scattering of low energy neutrons by ortho- and para-hydrogen.

(*d*) Diffraction of neutrons by solid hydrogen.

(*e*) Photo-electric cross-section for the disintegration of the deuteron by 2·62 MeV γ-rays.

(*f*) Proton-proton scattering.

(*g*) Triton binding energy.

Of these, the first five depend only on assumptions about the neutron-proton force. (*a*) may be somewhat unreliable in view of the uncertain position about the theory of the deuteron moment. In (*b*) it was assumed that the forces are of type I, which gives the best fit with the observed cross-sections. If type III forces were assumed, the best range would be slightly less but good agreement over the whole range of neutron energy could not be obtained with any range. For both, the ranges are compatible with the quadrupole moment of the deuteron.

In the determination of range from (*c*), (*d*) and (*e*) the interaction has been taken as purely central because the inclusion of the non-central part does not seriously affect the theoretical predictions (Hepner & Peierls 1942), provided the energies of the ^{1}S and ^{3}S states of the deuteron are given correctly. The ranges obtained from (*c*) and (*d*) and possibly from (*e*) are so small, however, that no choice of the constant γ in the non-central form (1) would give rise to the observed quadrupole moment. Thus in § 3·2 it was noted, by reference to figure 1, that this would be so if the range were $< 1{\cdot}3 \times 10^{-13}$ cm.

Proton-proton scattering gives, at present, information on the ^{1}S interaction between protons. The present experiments are still consistent with equality between this interaction and the corresponding one between neutron and proton if the range is taken to be the same. Assuming this to be true, the range which best fits the proton-proton scattering should be the same as that derived for neutron-proton interaction from other data.

The binding energy of the triton, though practically independent of the exchange character of the forces, may depend, for a given shape and range, on the non-central nature of the interaction. No detailed calculations have yet been carried out with

---

[*] A preliminary account has been given by the authors in a letter to *Nature* (Massey & Hu 1947).

TABLE 5. COMPARISON OF THE RANGES OF NUCLEAR FORCES FOR VARIOUS SHAPES OF INTERACTION DERIVED FROM DIFFERENT DATA

range in 10^{-13} cm.

	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)	(*f*)	(*g*)
shape of interaction	from the magnetic moment of the deuteron	from scattering of 1–6 MeV neutrons by protons	from scattering of thermal neutrons by hydrogen molecules	from diffraction of neutrons by hydrogen in crystals‡	from photo-disintegration of the deuteron by 2·62 MeV γ-rays	from scattering of protons by protons	from the triton binding energy
	Arnold & Roberts 1947	Bailey *et al.* 1946	Sutton *et al.* 1947	Shull *et al.*	—	Breit *et al.*	—
spherical well	2·8*	1·8–2·4	1·34–1·74	1·4 –1·8	0·7–1·3§ (1·7–2·2)	2·8	—
Gaussian	1·7–2·1†	—	—	—	—	1·94	2·25¶
Yukawa	1·3–1·5†	>1·2	0·75–1·06	0·65–1·1	0·4–0·7§ (1·1–1·4)	1·17–1·37‖	1·83**
exponential	1·5–1·7†	>0·95	0·70–0·98	0·65–1·0	0·5–0·8§ (0·9–1·3)	1·73	1·73††

* Rarita & Schwinger 1941.

† Estimated from the results given in table 2.

‡ The ranges given in this column should be the same as those given in the previous column if the experimental results used are sufficiently accurate.

§ These ranges were determined by assuming the photoelectric disintegration cross-section to be $7\cdot8 \times 10^{-28}$ cm.^{2} (Graham & Halban 1945). Recent measurements by Halban reported by Rosenfeld (1948) suggest that the cross-section should be at least 10×10^{-28} cm.^{2}. Support for this is provided by the recent measurements of Russell, Sachs, Wattenberg & Fields (1948) using 2·76 MeV γ-rays. Reducing their values to 2·62 MeV gives 9×10^{-28} cm.^{2} for the photoelectric cross-section. Values of the ranges given in brackets have been derived by assuming the cross-section to be between 9 and 10×10^{-28} cm.^{2}.

‖ The upper limit was determined by Lubanski & De Jager (Rosenfeld 1948).

¶ Margenau & Warren (1937), Margenau & Tyrrell (1938).

** Brown (1939).

†† Rarita & Present (1937).

such an interaction. The ranges given are those derived on the assumption of central forces and charge independence of the fundamental interactions.

Even ignoring the data of (*c*), (*d*) and (*e*) the agreement obtained between the ranges derived in the different ways is not good, for any shape of interaction, and is particularly bad for the Yukawa potential. The exponential potential gives the most consistent results.

A way out of these difficulties could, of course, be found by utilizing the patent remedy of theoretical nuclear physics, i.e. introducing more adjustable constants into the interaction potential. One of the obvious ways of doing this is to use different ranges for the ordinary, the spin exchange and the tensor forces, in other words we may assume both g and γ to be functions of the relative distance between the nucleons. It still remains to be seen how far this would have to be carried out.

The situation is clearly very fluid and we are yet far from a consistent picture of the nature of the fundamental interaction.

MATHEMATICAL APPENDIX

I. *Solution of the differential equations for the ground state of the deuteron*

a. General formulation of the iteration method

The usual task in a nuclear two-body problem is to determine the constants in the interaction potential with a given binding energy. For this purpose, it is more convenient to transform the Schrödinger equation into an integral equation and then solve it by an iteration method. The advantage of this procedure is that the interaction constants are determined in the process of calculation while, with the usual method, we have to determine them by trial and error.

We shall henceforth use the range r_0 of the force as the unit of length. The Schrödinger equation for the relative motion of two nucleons is then

$$\nabla^2\psi + (aU(\mathbf{r}) - \epsilon^2)\,\psi = 0, \tag{29}$$

where

$$\epsilon^2 = MEr_0^2/\hbar^2, \quad U(\mathbf{r}) = -\frac{Mr_0^2}{a\hbar^2}\mathscr{V}(\mathbf{r}).$$

The solution of (29), conforming to the boundary conditions of a bounded state, satisfies the integral equation

$$\psi(\mathbf{r}) = \frac{a}{4\pi}\int \frac{\exp[-\epsilon\,|\mathbf{r}-\mathbf{r}'|]}{|\mathbf{r}-\mathbf{r}'|}\,U(\mathbf{r}')\,\psi(\mathbf{r}')\,d\tau'. \tag{30}$$

To solve (30), a series of iterated functions

$$\phi_n(\mathbf{r}) = \frac{1}{4\pi}\int \frac{\exp[-\epsilon\,|\mathbf{r}-\mathbf{r}'|]}{|\mathbf{r}-\mathbf{r}'|}\,U(\mathbf{r}')\,\phi_{n-1}(\mathbf{r}')\,d\tau', \tag{31}$$

are constructed from a properly chosen initial function $\phi_0(\mathbf{r})$. Then the successive approximations to ψ and a are given by

$$a_n = \left\{\int|\phi_{n-1}|^2\,d\tau \Big/ \int|\phi_n|^2\,d\tau\right\}^{\frac{1}{2}}, \quad \psi_n = \phi_n \Big/ \left\{\int|\phi_n|^2\,d\tau\right\}^{\frac{1}{2}}. \tag{32}$$

It can be shown that the process does converge to the true solution of (30) (Svartholm 1945).

b. Application of the method to determine a and γ

For the interaction assumed in this paper

$$U(r) = (1+\gamma S_{12})\, V(r).$$

To ensure rapid convergence, ϕ_0 is calculated from the equation

$$\phi_0 = \frac{1}{4\pi}\int \frac{\exp[-\epsilon\,|\mathbf{r}-\mathbf{r}'|]}{|\mathbf{r}-\mathbf{r}'|}(1+\gamma S_{12})\, V(r')\, f_0(r')\,\chi_1 d\tau', \tag{33}$$

with
$$f_0(r) = \mathrm{e}^{-\epsilon}\frac{\sin kr}{r\sin k} \quad (r<1)$$
$$= r^{-1}\,\mathrm{e}^{-\epsilon r} \quad (r>1)$$

and
$$k\cot k = -\epsilon.$$

With ϕ_0 given by (33), it can be shown that the iterated functions $\phi_n(r)$ are all of the form

$$\phi_n = r^{-1}[f_{n,0}(r) + 8^{-\frac{1}{2}}S_{12}f_{n,2}(r)]\,\chi_1,$$

and (31) become simply (Watson 1944):

$$\left.\begin{aligned} f_{n0}(r) &= r^{\frac{1}{2}}I_{\frac{1}{2}}(\epsilon r)\int_r^\infty r'^{\frac{1}{2}}K_{\frac{1}{2}}(\epsilon r')\,[f_{n-1,0}(r') + \sqrt{8}\,\gamma f_{n-1,2}(r')]\,V(r')\,dr' \\ &\quad + r^{\frac{1}{2}}K_{\frac{1}{2}}(\epsilon r)\int_0^r r'^{\frac{1}{2}}I_{\frac{1}{2}}(\epsilon r')\,[f_{n-1,0}(r') + \sqrt{8}\,\gamma f_{n-1,2}(r')]\,V(r')\,dr', \\ f_{n2}(r) &= r^{\frac{1}{2}}I_{\frac{5}{2}}(\epsilon r)\int_r^\infty r'^{\frac{1}{2}}K_{\frac{5}{2}}(\epsilon r')\,[\sqrt{8}\,\gamma f_{n-1,0}(r') + (1-2\gamma)\,f_{n-1,2}(r')]\,V(r')\,dr' \\ &\quad + r^{\frac{1}{2}}K_{\frac{5}{2}}(\epsilon r)\int_0^r r'^{\frac{1}{2}}I_{\frac{5}{2}}(\epsilon r')\,[\sqrt{8}\,\gamma f_{n-1,0}(r') + (1-2\gamma)\,f_{n-1,2}(r')]\,V(r')\,dr', \end{aligned}\right\} \tag{34}$$

where I and K are Bessel functions of first and second kind with imaginary arguments. (The notation is that used by Watson 1944.)

The nth approximations for a and γ are given by the relations

$$a_n = \left[\int (f_{n,0}^2 + f_{n,2}^2)\,dr \Big/ \int (f_{n-1,0}^2 + f_{n-1,2}^2)\,dr\right]^{-\frac{1}{2}},$$

and
$$\frac{r_0^2}{10}\,\frac{\int r^2(\sqrt{2}\,f_{n,0}f_{n,2} - \frac{1}{2}f_{n,2}^2)\,dr}{\int (f_{n,0}^2 + f_{n,2}^2)\,dr} = 2{\cdot}73\times 10^{-27}\,\text{cm.}^2 \tag{35}$$

The second relation is the condition for the deuteron to have the correct electric quadrupole moment. It is to be noted that a simple justification of the general formulation (*a*) is possible only when γ is kept constant throughout the iteration processes. The procedure given here is reliable only when the variations of γ in the successive approximations are not very large.

In the case of the spherical well interaction, $V(r)$ vanishes outside the range of the force. Inside the range of the force, the interaction is much larger than the binding energy of the deuteron. To take the full advantage of these circumstances, a slightly modified procedure has been used. Referring to the differential equations (13), the solutions f_0 and f_2 now satisfy the integral equations

$$\left.\begin{aligned} f_0(r) &= (a-\epsilon^2)\left\{\int_0^r xf_0\,dx + r\int_r^1 f_0\,dx\right\} + 2\sqrt{2}\,\gamma a\left\{\int_0^r xf_2\,dx + r\int_r^1 f_2\,dx\right\} + c_1 r,\\ 5f_2(r) &= \sqrt{8}\,\gamma a\left\{r^{-2}\int_0^r x^3 f_0\,dx + r^3\int_r^1 x^{-2} f_0\,dx\right\}\\ &\qquad + \{a(1-2\gamma)-\epsilon^2\}\left\{r^{-2}\int_0^r x^3 f_2\,dx + r^3\int_r^1 x^{-2} f_2\,dx\right\} + c_2 r^3, \end{aligned}\right\} \tag*{36)}$$

for $r<1$. These functions must be joined smoothly to the solutions outside the range, viz.

$$f_0 = \mathrm{e}^{-\epsilon r}, \quad f_2 = q\,\mathrm{e}^{-\epsilon r}\left(\frac{1}{r^2}+\frac{\epsilon}{r}+\frac{\epsilon^2}{3}\right).$$

The conditions for a smooth joint at $r = 1$ are

$$\left.\begin{aligned} c_1 = -\epsilon\,\mathrm{e}^{-\epsilon}, \quad c_2 &= -\tfrac{1}{3}(\epsilon^2+\epsilon^3)\,q\,\mathrm{e}^{-\epsilon},\\ (1+\epsilon)\,\mathrm{e}^{-\epsilon} &= (a-\epsilon^2)\int_0^1 xf_0\,dx + \sqrt{8}\,\gamma a\int_0^1 xf_2\,dx,\\ q\,\mathrm{e}^{-\epsilon}\left[5\left(1+\epsilon+\frac{\epsilon^2}{3}\right)+\tfrac{1}{3}(\epsilon^2+\epsilon^3)\right] &= \sqrt{8}\,\gamma a\int_0^1 x^3 f_0\,dx + \{a(1-2\gamma)-\epsilon^2\}\int_0^1 x^3 f_2\,dx. \end{aligned}\right\} \tag{37}$$

The simultaneous integral equations (36) are again solved by an iteration method. The constants a, γ and q are determined for each approximation by the last two equations of (37) and the second equation of (35). The initial functions for the iteration were taken to be

$$f_0 = \mathrm{e}^{-\epsilon}\frac{\sin kr}{\sin k}, \quad f_2 = q\left(1+\epsilon+\frac{\epsilon^2}{3}\right)\mathrm{e}^{-\epsilon}\frac{\sin kr}{\sin k},$$

where k is given by $k\cot k = -\epsilon$.

As an illustration of the convergence of the method described above, the value of a and γ obtained by successive approximation for three different potential wells are given in table 6.

TABLE 6

type of potential well	a_1	a_2	a_3	γ_1	γ_2	γ_3
spherical well ($r_0 = 2{\cdot}62\times10^{-13}$ cm.)	1·95	1·93	—	1·26	1·29	—
Yukawa well ($r_0 = 1{\cdot}74\times10^{-13}$ cm.)	1·87	1·94	—	0·763	0·730	—
exponential well ($r_0 = 1{\cdot}74\times10^{-13}$ cm.)	5·64	6·09	6·20	0·94	0·85	0·832

II. *Calculation of the phases*

In the present work, over 500 phases had to be calculated. We shall give only a brief sketch of the methods used in these calculations.

The various continuous states can be divided into coupled and uncoupled states. The phases for a pair of coupled states are determined by the simultaneous differential equations (11) and the asymptotic conditions (19) and (22). For uncoupled states, a single differential equation and the usual asymptotic form $\sin(kr-\frac{1}{2}l\pi+\eta_l)$ is sufficient to determine the phase η_l.

a. The uncoupled states

All singlet states and the triplet states with $l = J$ are uncoupled states. To these one may add those states with $l = J-1$ when their coupling with states $l = J+1$ can be neglected. Within the energy range considered in this paper, the coupling of $^3P_2$ and $^3F_2$ states can be neglected in the case of a spherical well interaction since the F-wave does not rise to sufficient amplitude within the range of force to perturb seriously the P-wave coupled with it. For Yukawa and exponential wells, the couplings between $^3F_4$ and $^3H_4$, and $^3G_5$ and $^3I_5$ states can be neglected.

The differential equation for an uncoupled state is

$$\left(\frac{d^2}{dx^2}-\frac{l(l+1)}{x^2}+k^2\right)f_l(x) = -aV(x)f_l(x). \tag{38}$$

For the spherical well case, the phase η_l is given by

$$\tan\eta_l = (-1)^l A_l/B_l,$$
$$A_l = kg_{l-1}(k)\,g_l(k') - k'g_{l-1}(k')\,g_l(k),$$
$$B_l = kg_{-l+1}(k)\,g_l(k') + k'g_{l-1}(k')\,g_{-l}(k),$$

where $$k' = (a+k^2)^{\frac{1}{2}}, \quad g_l(x) = \left(\frac{\pi x}{2}\right)^{\frac{1}{2}} J_{l+\frac{1}{2}}(x).$$

For other potentials η_0 was calculated by numerical integration of (37), and, for $l \geqslant 2$, η_l was calculated by Born's approximation

$$\eta_l(B) = \frac{a}{k}\int_0^\infty \{g_l(kx)\}^2\,V(x)\,dx. \tag{39}$$

To calculate η_1, the simplest version of Hulthén's method (Hulthén 1944) has been used. As is well known, the phase η_l is related to the wave function f_l by the formula

$$\sin\eta_l = \frac{a}{k}\int_0^\infty f_l\,g_l\,V(x)\,dx. \tag{40}$$

If we replace f_l by g_l we obtain (39). Writing f_l now in the form

$$f_l = g_l(kx)\cos\eta_l + h_l(x)\sin\eta_l. \tag{41}$$

Substituting in (40), and solving for η_l, gives

$$\cot\eta_l = [\eta_l(B)]^{-1}\left\{1-\frac{a}{k}\int_0^\infty g_l h_l\,V(x)\,dx\right\}. \tag{42}$$

According to Hulthén, $h_l(x)$ can be put approximately in the form

$$(-1)^l\,(1+be^{-x})\,(1-e^{-x})^{2l+1}\,g_{-l}(kx).^*$$

The value of b is determined by the condition

$$\int_0^\infty f_l\left[\frac{d^2}{dx^2}-\frac{l(l+1)}{x^2}+k^2+aV(x)\right]f_l\,dx = 0. \tag{43}$$

With the help of (42), η_l can be eliminated from (43). We get, then, a quadratic equation for b. To select one of the roots, a further condition

$$\frac{\partial}{\partial b}\int_0^\infty f_l\left[\frac{d^2}{dx^2}-\frac{l(l+1)}{x^2}+k^2+aV(x)\right]f_l\,dx = 0 \tag{44}$$

---

* This will be a poor approximation when the first zero of $g_{-l}(kx)$ lies within the range of force. (This will not happen in all cases considered in this paper.)

has been used. The adequacy of this method can also be tested by this relation. As a further check to this method, numerical integrations have been performed for several cases. The results are given in table 7.

TABLE 7

For Yukawa potential $V(x) = e^{-x}/x$, $k = 1{\cdot}062$

a	4·772	−4·772	−1·591
η_1 by the approximate method	0·7983	−0·3959	−0·1564
η_1 by numerical integration	0·7974	−0·3961	−0·1557

For exponential well $V(x) = e^{-2x}$, $k = 1{\cdot}062$

a	16·52	−16·52	−5·504
η_1 by the approximate method	0·9642	−0·3520	−0·1470
η_1 by numerical integration	1·0161	−0·3510	−0·1466

It is of interest to notice that an interpolation formula for η_l can be derived from (42) by expanding $h(x)$ as a power series in a. We get, then,

$$(\eta_l(B)\cot\eta_l - 1) = c_1 a + c_2 a^2 + c_3 a^3 + \ldots, \tag{45}$$

where c_1, c_2, etc. are coefficients depending on the value of k only.

With two or three coefficients determined by known values of η_l (45) represents a quite accurate relation between a and η_l. We have used this formula to check our numerical calculations.

b. The coupled states

For the coupled states $^3S_1$ and $^3D_1$, in the case of the spherical well, the differential equations (11) were solved by series expansion. For other shapes of potential, direct numerical integrations of the simultaneous differential equations was employed. (We are indebted to the Scientific Computing Service for this numerical work.)

For other states, the following approximate methods were used. The wave functions $f_{J-1}(x)$, $f_{J+1}(x)$ are connected with the phases by the relations

$$\left.\begin{aligned} \sin\eta_{J-1,M} &= \frac{1}{k\kappa_{J-1,M}}\int_0^\infty g_{J-1}(kx)\,V(x)\,[af_{J-1}(x)+bf_{J+1}(x)]\,dx,\\ \sin\eta_{J+1,M} &= \frac{1}{k\kappa_{J+1,M}}\int_0^\infty g_{J+1}(kx)\,V(x)\,[bf_{J-1}(x)+cf_{J+1}(x)]\,dx,\end{aligned}\right\} \tag{46}$$

where a, b and c are the coefficients of $V(x)$ in (11) and $\kappa_{J\pm1,M}$ are as defined in (22). For Born's approximation, we put $f_{J\pm1}(x) = \kappa_{M,J\pm1}g_{J\pm1}(kx)$ and neglect all higher orders of η. We get, then,

$$\left.\begin{aligned} \eta_{J-1,M} &= \frac{a}{k}\int_0^\infty \{g_{J-1}(kx)\}^2\,V(x)\,dx + \frac{b}{k}\frac{\kappa_{J+1,M}}{\kappa_{J-1,M}}\int_0^\infty g_{J-1}g_{J+1}\,V(x)\,dx,\\ \eta_{J+1,M} &= \frac{b}{k}\frac{\kappa_{J-1,M}}{\kappa_{J+1,M}}\int_0^\infty g_{J+1}g_{J-1}\,V(x)\,dx + \frac{c}{k}\int_0^\infty \{g_{J+1}(kx)\}^2\,V(x)\,dx.\end{aligned}\right\} \tag{47}$$

In the case of $J = 2$, a modified Born's approximation was used. The improvement consisted simply in replacing $f_{J\pm1}$ by

$$\left.\begin{aligned} f_{J-1}(x) &= \kappa_{J-1,M}[g_{J-1}(kx)\cos\eta_{J-1,M} + (-1)^{J-1}\sin\eta_{J-1,M}\, g_{-J+1}(kx)\,(1-\mathrm{e}^{-x})^{2J-1}],\\ f_{J+1}(x) &= \kappa_{J+1,M}[g_{J+1}(kx)\cos\eta_{J+1,M} + (-1)^{J+1}\sin\eta_{J+1,M}\, g_{-J-1}(kx)\,(1-\mathrm{e}^{-x})^{2J+2}].\end{aligned}\right\} \tag{48}$$

Substituting (48) into (46), we obtain a pair of simultaneous equations from which η_{J+1} and η_{J-1} can be determined. As a check of the accuracy of the method, numerical integration has been performed for a single case with a Yukawa well interaction. The results, together with those obtained from both Born's approximation and the modified version are given in table 8.

TABLE 8

$\eta_{lM} = \lambda_{lM} + i\zeta_{lM}$, $k = 1{\cdot}062$

$a = 1{\cdot}034$, $b = 3{\cdot}791$ and $c = -0{\cdot}3134$

	λ_{10}	λ_{11}	λ_{30}	λ_{31}
numerical integration	0·3176	0·1783	0·0487	−0·0897
modified Born's approx.	0·3068	0·1567	0·0526	−0·0883
Born's approx.	0·243	0·088	0·062	−0·093
	ζ_{10}	ζ_{11}	ζ_{30}	ζ_{31}
numerical integration	−0·0225	0·0161	0·0161	−0·0223
modified Born's approx.	−0·0228	0·0158	+0·0154	−0·0207
Born's approx.	—	—	—	—

We are indebted to the Government Grant Committee of the Royal Society for a grant which enabled us to obtain the assistance of the Scientific Computing Service who provided the numerical solution of certain coupled equations occurring with the exponential and Yukawa interactions.

It is a pleasure to thank the British Council for the award of a grant which enabled one of us (T.H.) to spend the time in this country during which this work was carried out.

We also wish to thank Dr Burhop for suggestions concerning the use of photo-disintegration in determining the range of the force, and Dr Buckingham for the discussion of points arising in the numerical work.

REFERENCES

Amaldi, E., Bocciarelli, D., Ferretti, B. & Trabacchi, G. C. 1942 *Naturwissenschaften*, **30**, 582.
Amaldi, E., Ageno, M., Bocciarelli, D. & Trabacchi, G. C. 1943 *Naturwissenschaften*, **31**, 231.
Arnold, W. R. & Roberts, A. 1947 *Phys. Rev.* **71**, 878.
Bailey, C. E., Bennett, W. E., Bergstrahl, T., Nuckolls, R. G., Richards, H. T. & Williams, J. H. 1946 *Phys. Rev.* **70**, 583.
Bethe, H. A. 1940 *Phys. Rev.* **57**, 260 and 390.
Breit, G. & Bloch, J. 1947 *Phys. Rev.* **72**, 135.
Brown, F. W. 1939 *Phys. Rev.* **56**, 1107.
Fröhlich, H., Ramsey, W. H. & Sneddon, I. N. 1947 *Report on an international conference on fundamental particles and low temperatures*, Vol. I, p. 166. London: Physical Society.

Graham, G. A. R. & Halban, H. (Jr) 1945 *Rev. Mod. Phys.* **17**, 297.
Halban, H. (Jr) 1938 *C.R. Acad. Sci., Paris*, **206**, 1170.
Hepner, W. & Peierls, R. 1942 *Proc. Roy. Soc.* A, **181**, 43.
Hoisington, L. E., Share, S. S. & Breit, G. 1939 *Phys. Rev.* **56**, 884.
Hulthén, L. 1943 *Ark. Mat. Ast. Fys.* **29**A, no. 33 and **30**A, no. 9.
Hulthén, L. 1944 *K. fysiogr. Sällsk. Lund. Förh.* **14**, 257.
Jauch, J. M. 1945 *Phys. Rev.* **68**, 125.
Kellogg, J. M. B., Rabi, I. I., Zacharias, J. R. & Ramsey, N. F. 1940 *Phys. Rev.* **57**, 677.
Kittel, C. & Breit, G. 1939 *Phys. Rev.* **56**, 744.
Laughlin, J. S. & Kruger, P. G. 1948 *Phys. Rev.* **73**, 197.
Margenau, H. & Tyrrell, W. A. (Jr) 1938 *Phys. Rev.* **54**, 422.
Margenau, H. & Warren, D. T. 1937 *Phys. Rev.* **52**, 790, 1027.
Massey, H. S. W. & Buckingham, R. A. 1937 *Proc. Roy. Soc.* A, **163**, 281.
Massey, H. S. W. & Hu, T. M. 1947 *Nature*, **160**, 794.
Millman, S. & Kusch, R. 1941 *Phys. Rev.* **60**, 91.
Møller, C. & Rosenfeld, L. 1940 *Math.-fys. Medd.* **17**, no. 8.
Pais, A. 1946 *Proc. Camb. Phil. Soc.* **42**, 45.
Powell, C. F. & Occhialini, G. P. S. 1947 *Report of an international conference on fundamental particles and low temperature*, Vol. 1, p. 150. London: Physical Society.
Primakoff, H. 1947 *Phys. Rev.* **72**, 118.
Ramsey, W. H. 1947 *Proc. Roy. Soc.* A, **191**, 195.
Rarita, W. & Present, R. 1937 *Phys. Rev.* **51**, 788.
Rarita, W. & Schwinger, J. 1941 *Phys. Rev.* **59**, 436, 556.
Rosenfeld, L. 1948 *Nuclear forces*. Amsterdam: North Holland Publishing Company.
Russell, B., Sachs, D., Wattenberg, A. & Fields, R. 1948 *Phys. Rev.* **73**, 545.
Sachs, R. G. 1947 *Phys. Rev.* **72**, 91.
Salant, E. O. & Ramsey, N. F. 1940 *Phys. Rev.* **57**, 1075.
Sherr, R. 1945 *Phys. Rev.* **68**, 240.
Shull, G. C., Wollan, E. D., Morton, G. A. & Davidson, W. L. 1948 *Phys. Rev.* **73**, 842.
Sleater, W. (Jr) 1947 *Phys. Rev.* **72**, 207.
Sutton, R. B., Hall, T., Anderson, E. E., Bridge, H. S., De Wire, J. H., Lavatelli, L. S., Long, E. A., Snyder, T. & Williams, R. W. 1947 *Phys. Rev.* **72**, 1147.
Svartholm, N. 1945 *The binding energies of the lighest atomic nuclei*. Lund: Hakan Ohlsson.
Villars, F. 1947 *Helv. Phys. Acta*, **20**, 476.
Volkhoff, G. M. 1942 *Phys. Rev.* **62**, 126.
Watson, G. N. 1944 *A treatise on the theory of Bessel functions*, p. 366. Cambridge University Press.

Influence of chlorine substitution on the rate of oxidation of paraffin hydrocarbons

By C. F. Cullis, Sir Cyril Hinshelwood, F.R.S. and M. F. R. Mulcahy*
Physical Chemistry Laboratory, University of Oxford

(*Received* 23 *July* 1948)

In the kinetics of their gas phase oxidations primary and some secondary chloro-paraffins are generally similar to the corresponding hydrocarbons. The rate is strongly affected by the pressure of the organic reactant and nearly independent of oxygen. (With tertiary chloro-compounds an oxygen-catalyzed pyrolysis obscures the direct oxidation.)

For the oxidation a mechanism analogous to that proposed for hydrocarbons is assumed, namely one in which the rate is governed by branching chains whose development depends upon the breaking of an unstable —O—O— link.

Previous work on the influence of hydrocarbon structure is interpreted with the aid of the hypothesis that methyl groups exert a specific stabilizing effect (and that relatively rapid oxidation demands a point of attack in the molecule sufficiently remote from a methyl group).

In the light of this hypothesis the influence of chlorine atoms is compared with that of methyl groups. The chlorine accelerates the oxidation (*a*) by impairing the symmetry of a methyl group into which it is introduced, and (*b*) by a direct inductive effect of its own.

The opposite effects of chlorine and of methyl are consistent with the view that the —O—O— links are stabilized by electron accession and weakened by electron recession.

INTRODUCTION

One of the most remarkable characteristics of hydrocarbon oxidation is the profound influence of structure on reaction rate. It expresses itself both in a rapid increase in velocity as a normal paraffin chain is lengthened, and in an equally striking decrease as a molecule of given carbon content becomes more branched. These effects are specific to the oxidation reaction and do not affect the pyrolysis in anything like a comparable way. The object of the present paper is to consider the influence of chlorine substitution upon oxidation.

Here we have to deal with the effect of a given substituent atom, and it will facilitate subsequent discussion to explain first that in the light of the evidence now available we consider the influence of structure on the oxidation of normal and branched paraffins to be very largely a function of the methyl groups in the molecule. The present point of view may be stated as follows. Methyl groups appear in the first place to be inherently resistant to initial attack (compare the very low oxidation rates of methane and ethane), a matter which is being considered further in another paper. Normally, therefore, the initial attack of oxygen on the hydrocarbon is at a CH_2 or CH group. In this respect the behaviour of oxygen is similar to that of other electrophilic reagents, e.g. alkyl radicals and Cl atoms (Smith & Taylor 1939; Rice & Rice 1935; Steiner & Watson 1947). Methyl groups also exert a specific effect in hindering the development of the branching chain (which must be postulated to explain the kinetic phenomena as a whole). It is the development rather than the initiation of the branching chains which will have the greatest influence on the rate.

---

[*] Member of the staff of the Division of Tribophysics, C.S.I.R., Australia.

We assume, indeed, that the molecules which contribute to the reaction are overwhelmingly those in which the initial attack has been suitable for this rapid development. (Even if another initial mode of attack were somewhat easier it would contribute little, because the chain-branching which ensues from it is relatively so much less important.)

Without for the moment making any further hypothesis about the nature of the stabilizing effect of methyl groups the assumption of its existence suffices to correlate the structural influences just referred to, which are otherwise rather difficult to bring under one head.

We assume the oxidation rate to be some inverse function of a 'stability factor', calculated by assigning unit contribution to a methyl group adjacent to that centre in the molecule where the initial oxygen attack has occurred, 1/3 to a group one atom removed and so on. The decrement of 1/3 is taken by analogy with the work of Derick (1911) on inductive effects in acid dissociation.

The stability factor, S, for each possible reaction centre of a given paraffin is calculated and the chain ramification taken to be determined by the lowest possible value of S. A graph of S against the logarithm of the reaction rate shows a single functional relation for normal and branched paraffins, as may be seen in figure 6.

According to this way of regarding the matter long normal paraffins oxidize readily because they contain points of attack far enough removed from CH_3 groups, while branched structures are stable because they do not.

The effect of the methyl group may reasonably be ascribed to its efficiency as an electron source, and this in turn may very possibly be due to its symmetry and its aptitude for participation in electron delocalization phenomena (Coulson 1947; Hinshelwood 1948). If the three hydrogens of the methyl group all contribute electron density to the fourth bond, then the transmission of negative charge to the rest of the molecule is facilitated. It is in this connexion that a comparison between the effects of methyl and of chlorine on hydrocarbon oxidation assumes special interest. The same phenomenon manifesting itself in a different way leads to a stabilization of the methyl group, which may, therefore, be expected to resist initial attack by oxygen.

For the purposes of the comparison between methyl and chlorine the interpretation of the reaction mechanisms involved in hydrocarbon oxidation need not be carried further. A more specific hypothesis, however, is relevant to the discussion of the general kinetic behaviour of the chloro-derivatives. Our working hypothesis about chain initiation and development has been discussed elsewhere (Cullis, Hinshelwood, Mulcahy & Partington 1947). The essentials are as follows. Initiation is by the steps

$$\begin{aligned} RH + O_2 &= R + HO_2 \\ R + O_2 &= R\text{—}O\text{—}O\text{—} \\ R\text{—}O\text{—}O\text{—} + RH &= R\text{—}O\text{—}O\text{—}R \\ R\text{—}O\text{—}O\text{—}R &= 2R\text{—}O\text{—} \end{aligned}$$

Alkoxyl radicals of the right type break down to give lower alkyl radicals + aldehydes. The former may react with oxygen to suffer further degradation or with paraffin to regenerate R.

The stabilizing effect of electron sources is assumed to consist in a strengthening of the peroxidic link, and the consequent hindrance of chain development.

The most striking kinetic characteristics which this scheme summarizes are:

(*a*) the existence in the low temperature oxidation of an induction period and the gradual acceleration of the reaction to a maximum rate;

(*b*) independence, over a considerable range, of oxygen pressure and reaction rate;

(*c*) steep increase of rate with concentration of hydrocarbon.

If the oxidation of the chlorine-substituted hydrocarbons takes place according to a mechanism basically similar to that which obtains for the unsubstituted compounds, the same kinetic characteristics should again be in evidence.

In what follows the problems to be considered are:

(1) To what extent the reactions of the chloro derivatives are comparable in mechanism with those of the simple paraffins.

(2) Where the mechanisms are in fact comparable, what are the relative influences of methyl and chlorine.

EXPERIMENTAL

Methods

The vapour of chloro-hydrocarbon and oxygen were admitted separately to a closed silica reaction vessel at constant temperature and the reaction followed manometrically. Details, including methods of analysis, were as previously described (compare Cullis *et al.* 1947). The reaction is very sensitive to surface conditions. At first erratic results are obtained, but after a number of preliminary runs in the vessel with standardization of evacuation procedure reasonably consistent values for the rate are achieved. Two separate reaction vessels of 290 and 270 ml. respectively were used, the rates in the larger vessel being the greater. A certain irreproducibility is almost certainly due to variations in the chain-controlling character of the vessel walls, but it is not serious enough to mask the broad structural comparisons which are to be considered.

The pressure-time curves

With most of the chloro-paraffins a gradual build-up in the reaction rate was observed, the pressure-time curves being, in a general way, analogous to those obtained with the unsubstituted substances.

The effect of reactant concentrations on the reaction rate

This was investigated by measuring at constant temperature the maximum rates of oxidation ($\rho_{\text{max.}}$) for various mixtures of chloro-paraffin and oxygen. Results are given below for 1-chloro-propane, 1-chloro-3-methyl-butane, 2-chloro-propane and 2-chloro-butane, the first being a primary substituted straight chain, the second a primary substituted branched chain and the third and fourth secondary substituted straight chains. It will be shown later that tertiary substituted compounds are unsuitable for investigation.

1-*Chloro-propane*. Figure 1 (curve *A*) refers to mixtures of 102 mm. initial oxygen pressure with various initial pressures of 1-chloro-propane at 271° C. The rate rises with rather more than the second power of the chloro-hydrocarbon pressure. Figure 2 (curve *a*), which refers to 30 mm. 1-chloro-propane with various pressures of oxygen at 271° C, shows the rate to be independent of oxygen concentration from 500 to 20 mm. initial pressure. Between 20 and 7 mm. the rate falls off rapidly.* It is evident

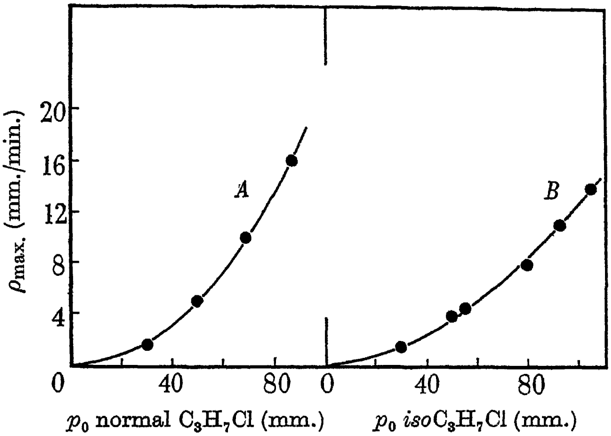


FIGURE 1. Curve *A*, oxidation of 1-chloro-propane at 271° C (O_2 = 102 mm.). Curve *B*, oxidation of 2-chloro-propane at 343·5° C (O_2 = 250 mm.).

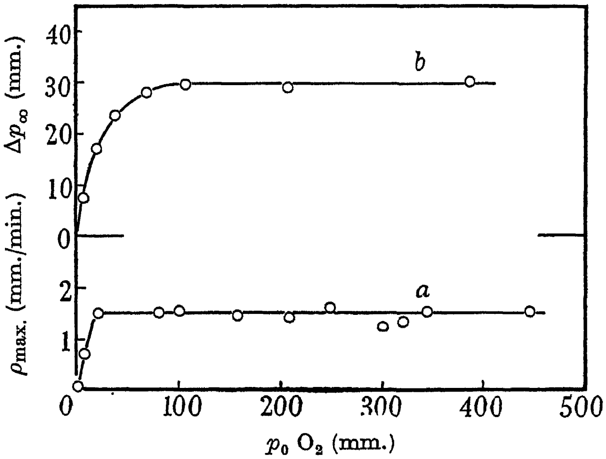


FIGURE 2. Rates ($\rho_{\max.}$) curve (*a*), and total pressure changes (Δp_{∞}) curve (*b*), for 1-chloro-propane (30 mm.) at 271° C.

that in the strong dependence of the rate on hydrocarbon concentration and the absence of any effect of oxygen over a wide range, the behaviour of this substance is similar to that of unsubstituted hydrocarbons.

Figure 2 (curve *b*) gives the final pressure increase on complete reaction (Δp_{∞}) for various pressures of oxygen and 30 mm. 1-chloro-propane. Δp_{∞} increases to a maximum value at about 100 mm. initial oxygen pressure, thereafter remaining substantially constant. Analogous behaviour is shown by unsubstituted paraffins (Cullis *et al.* 1947). Above 294° C (30 mm. C_3H_7Cl/250 mm. O_2) the reaction gives rise to a cool flame, a further indication that the general nature of the reaction is similar to that of unsubstituted hydrocarbons.

---

* The rate of pyrolysis in absence of oxygen is appreciable (0·05 mm./min.) under these conditions, though small compared with the rate of oxidation. There was no sign of an induction period in the absence of oxygen but with 7 mm. oxygen the reaction took 12 min. to attain its $\rho_{\max.}$ of 0·7 mm./min.

1-*Chloro*-3-*methyl-butane*. Figure 3 shows that the behaviour of this compound at 227° C is also analogous to that of the paraffins.

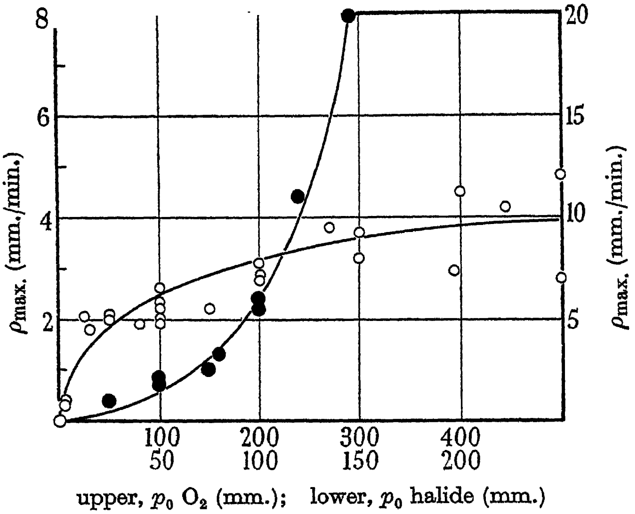


FIGURE 3. Oxidation of 1-chloro-3-methyl-butane at 227° C. Open circles, halide 50 mm., O_2 varied, left-hand scale. Filled-in circles, O_2 200 mm., halide varied, right-hand scale.

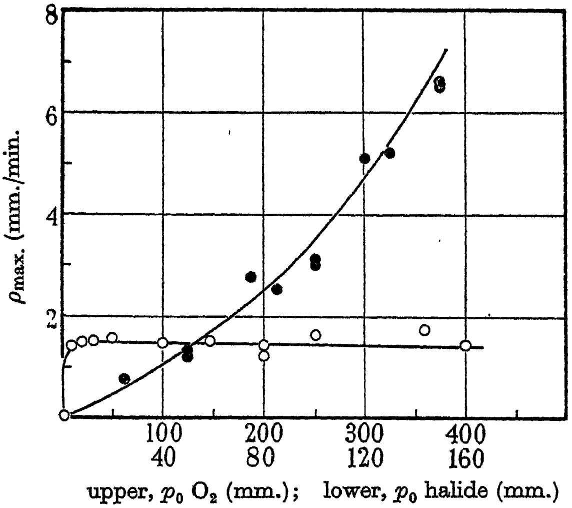


FIGURE 4. Oxidation of 2-chloro-butane at 220° C. Open circles, halide 50 mm., O_2 varied. Filled-in circles, O_2 200 mm., halide varied.

2-*Chloro-butane*. Figure 4 shows the behaviour at 220° C of this secondary compound. Dependence of $\rho_{max.}$ on the reactant concentrations is similar to that of the primary and of the unsubstituted compounds. The rate remains independent of oxygen concentration down to an initial oxygen pressure of less than 10 mm.

2-*Chloro-propane.* Figure 1 (curve *B*) shows the effect on $\rho_{\text{max.}}$ of increasing initial pressures of this compound at 343·5° C. Comparison with curve *A* shows that the rates are less than those found with 1-chloro-propane in spite of the 70° higher temperature. Figure 5 (curve *a*) shows that $\rho_{\text{max.}}$ is independent of oxygen concentration above 8 mm. The initial (maximum) rate of pyrolysis (50 mm. 2-chloro-propane) was 0·27 mm./min.

The evidence so far indicates similarity between the reactions of the primary and secondary substituted propanes. Nevertheless, the case is quite otherwise, as may be seen by comparison of the effects of initial oxygen concentrations on the final pressures reached (figures 2 (curve *b*) and 5 (curve *b*)). With 2-chloro-propane, Δp_∞ is almost independent of the oxygen pressure: there is no falling off of the pressure of products as the amount of oxygen is decreased such as is observed with the primary compound. In fact Δp_∞ reaches the full pyrolysis value of 50 mm. with only 0·2 mm. oxygen present. With increasing oxygen concentration, Δp_∞ decreases somewhat, reaching an apparently constant value of 46 mm. at about 250 mm. initial oxygen pressure.* The contrast with the behaviour of the primary compound is remarkable.

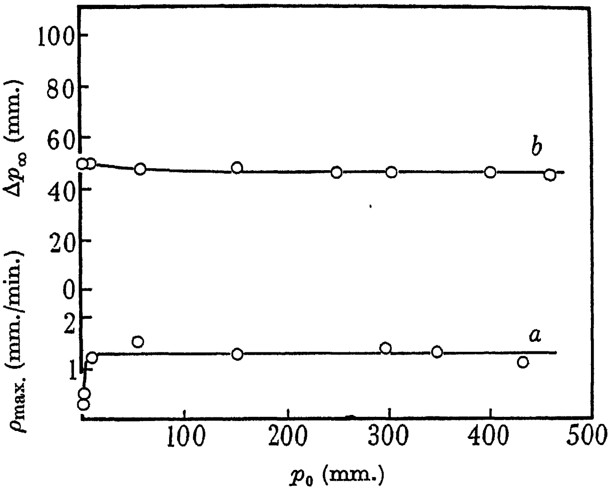


FIGURE 5. Oxidation of 2-chloro-propane at 343·5° C. Halide 50 mm., O_2 varied. Curve (*a*) max. rate. Curve (*b*) total pressure change, Δp_∞ (mm.)

It is apparent from this result that the main function of the oxygen is catalytic. Presumably the reaction observed is the pyrolysis to propylene and HCl catalyzed by oxygen. This view is supported by various experiments of which the following, with 30 mm. 2-chloro-propane and 250 mm. oxygen at 330° C, may be quoted as an example. The initial reaction was almost complete in 2 hr., but after a further 40 min. a fast reaction suddenly set in, which was due undoubtedly to oxidation of the products of the first stage.

* With 250 mm. oxygen and various pressures of 2-chloro-propane Δp_∞ is a little less than the original chloro-propane pressure, but increases linearly with it up to at least 130 mm. initial chloro-propane pressure, i.e. to over one-half the pressure of oxygen present.

The experiments with 2-chloro-propane clearly provide no values for the rate of oxidation suitable for quantitative comparison with 1-chloro-propane and other compounds. The rate of pyrolysis observed does, however, give an upper limit to the rate of oxidation and there is no doubt that the 2-compound oxidizes much less rapidly than the primary one.

Other secondary and tertiary compounds. The question arises whether other secondary (and tertiary) compounds are oxidized or pyrolyzed under the conditions of the experiments. With 30 mm. 2-chloro-butane and 250 mm. oxygen at 260·7° C. Δp_∞ was found to be 32 mm.; with 5 mm. oxygen Δp_∞ was 5 mm. On the other hand, 30 mm. of the isomeric tertiary compound 2-methyl-2-chloro-propane at the same temperature gave 28 mm. for Δp_∞ with 5 mm. oxygen which is nearly equal to the 34 mm. obtained with 250 mm. Moreover, the maximum (initial) rate of pyrolysis of the tertiary compound in the absence of oxygen at 260·7° C was found to be half the maximum rate of oxidation. A similarly high rate of pyrolysis was observed with 2-methyl-2-chloro-butane. It may be concluded that the reaction with the secondary substituted butane, and probably other secondary compounds, except 2-chloro-propane, is oxidation, but catalyzed pyrolysis occurs with the tertiary compounds.

Most of the substances for which the comparative rates have been measured are primary compounds. Apart from 1-chloro-propane (which is the least susceptible to oxidation), the dependence of the final pressure on oxygen pressure was not investigated. In view, however, of the comparative stability of the compounds towards pyrolysis, the result obtained with 1-chloro-propane, the similarity of the formal kinetics to those of the unsubstituted substances and the general structural consistency of the comparative oxidation rates (see below), there can be little doubt that the reaction of the compounds with oxygen is basically similar in nature to that which occurs with the unsubstituted paraffins.

Analytical results

Various analytical measurements were made by the methods stated in the previous paper (Cullis *et al.* 1947) on 1-chloro-pentane for comparison with those previously made for the paraffin oxidation. Both peroxides and aldehydes were detected. In the early stages of the reaction, the concentrations of peroxides and aldehydes run roughly parallel with the rate of pressure increase, and reach maximum values when the rate acquires its maximum value ($\rho_{\text{max.}}$). The results given in table 1 represent the maximum amounts of intermediates present at any stage of the reaction. On the assumption that the quantities of intermediates present are approximately proportional to the rate of pressure increase at that time, the numbers given have been reduced to equal values of $\rho_{\text{max.}}$. Peroxides are formed in roughly equal amounts in the two cases, but whereas in the oxidation of n-hexane approximately equivalent quantities of higher aldehydes and formaldehyde are formed, the aldehyde detected during the slow combustion of 1-chloro-pentane consists almost exclusively of formaldehyde. The reason for this difference is not yet clear. The chloro-aldehydes, which would presumably be formed from the latter compound, may polymerize or decompose rapidly under the experimental conditions, or, possibly, such compounds may not respond as readily as unsubstituted aldehydes to the methods of analysis employed.

TABLE 1. RELATIVE QUANTITIES OF INTERMEDIATE PRODUCTS FORMED DURING THE SLOW OXIDATION OF *n*-HEXANE AND OF 1-CHLORO-PENTANE

compound	*n*-hexane	1-chloro-pentane
peroxides (max.)	2·88	3·47
total aldehydes (max.)	5·55	1·83
formaldehyde (max.)	2·04	1·64

Discussion of structural influences

Hydrocarbons. In table 2 are shown the relative oxidation rates of a number of straight-chain and branched-chain hydrocarbons together with the minimum values of the stability factor, S, which have been calculated as explained in the introduction. As far as possible, the comparisons of oxidation rates have been made under identical conditions of concentration and temperature but, where this was not practicable because of the enormous differences in oxidizability between the compounds studied, measurements had to be made in pairs at a series of temperatures. Figure 6 shows the relationship between $\log_{10}$ (rate) and S, to which reference has already been made.

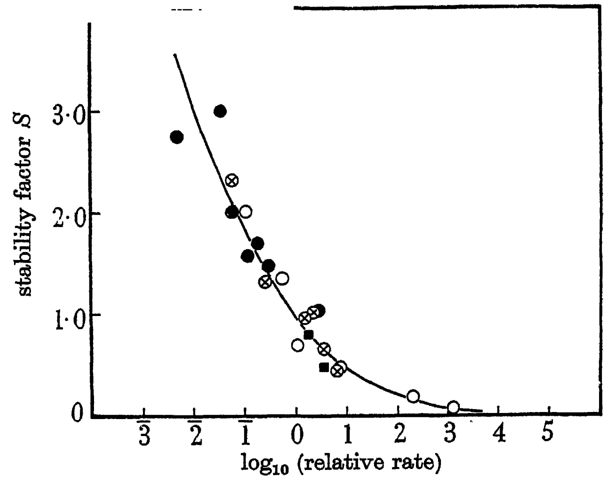


FIGURE 6. Stability factor and log (rate relative to pentane) for paraffins and chloro-paraffins. ○ straight-chain hydrocarbons; ⊗ primary chloro-hydrocarbons; ● branched-chain hydrocarbons; ■ secondary chloro-hydrocarbons.

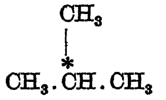
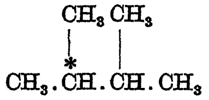
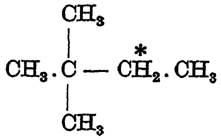
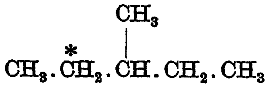
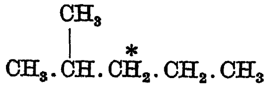
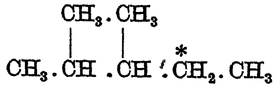
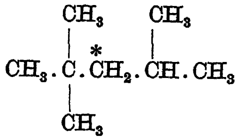
Primary chloro-paraffins. Table 3 (*a*) shows the relative oxidation rates and the stability factors for a number of primary chloro-paraffins. S was calculated in the following way.

The important point of attack of a hydrocarbon molecule is where the combined influence of the methyl groups is least. The group $—CH_2Cl$ may be regarded as a substituted methyl group, the symmetry of which has been impaired by the replacement of a hydrogen atom by a chlorine atom. Initial attack will be taken not to occur readily at this group. In the compound 1-chloro-butane,

$$CH_3.\overset{(1)}{C}H_2.\overset{(2)}{C}H_2.CH_2Cl,$$

the point of attack will probably be either at (1) or (2), but (2) is more likely than (1) since it is further from the unsubstituted methyl group, which has a more powerful stabilizing influence than the substituted group. If we assign to the group $—CH_2Cl$

TABLE 2. THE INFLUENCE OF STRUCTURE ON THE RATE OF OXIDATION OF SOME PARAFFIN HYDROCARBONS

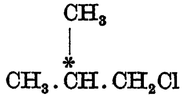
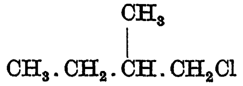
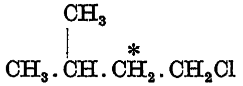
structure	oxidation rate (relative to *n*-pentane)	stability factor
(*a*) straight-chain hydrocarbons		
$CH_3.\overset{*}{C}H_2.CH_3$	0·1	2·0
$CH_3.\overset{*}{C}H_2.CH_2.CH_3$	0·5	1·33
$CH_3.CH_2.\overset{*}{C}H_2.CH_2.CH_3$	(1·0)	0·67
$CH_3.CH_2.\overset{*}{C}H_2.(CH_2)_2.CH_3$	7·5	0·44
$CH_3.(CH_2)_2.\overset{*}{C}H_2.(CH_2)_3.CH_3$	200	0·14
$CH_3.(CH_2)_3.\overset{*}{C}H_2.(CH_2)_4.CH_3$	1380	0·04
(*b*) branched-chain hydrocarbons		
 $CH_3.\overset{CH_3}{\overset{\vert}{\overset{*}{C}H}}.CH_3$	0·036	3·0
 $CH_3.\overset{CH_3}{\overset{\vert}{\overset{*}{C}H}}.\overset{CH_3}{\overset{\vert}{C}H}.CH_3$	0·0046	2·67
 $CH_3.\underset{CH_3}{\underset{\vert}{\overset{CH_3}{\overset{\vert}{C}}}}—\overset{*}{C}H_2.CH_3$	0·058	2·0
 $CH_3.\overset{*}{C}H_2.\overset{CH_3}{\overset{\vert}{C}H}.CH_2.CH_3$	0·28	1·44
 $CH_3.\overset{CH_3}{\overset{\vert}{C}H}.\overset{*}{C}H_2.CH_2.CH_3$	2·65	1·0
 $CH_3.\overset{CH_3}{\overset{\vert}{C}H}.\overset{CH_3}{\overset{\vert}{C}H}.\overset{*}{C}H_2.CH_3$	0·11	1·55
 $CH_3.\underset{CH_3}{\underset{\vert}{\overset{CH_3}{\overset{\vert}{C}}}}.\overset{*}{C}H_2.\overset{CH_3}{\overset{\vert}{C}H}.CH_3$	0·18	1·67

* Indicates assumed point of attack.

as a whole a contribution x (relative to $CH_3 = 1{\cdot}0$), the stability factor of the above compound will be: $x+\frac{1}{3}(1{\cdot}0)$. If, moreover, the substituted compounds are to show general conformity with the hydrocarbons themselves, then from the results for the latter (compare figure 6), 1-chloro-butane (for which $\log_{10}(\text{rate}) = 0{\cdot}300$) should have a stability factor of 0·66. Equating these values, $x+0{\cdot}33 = 0{\cdot}66$, whence $x = 0{\cdot}33$.

S may now be calculated for other primary chloro-compounds with the assumption that this contribution of the —CH_2Cl group is decreased to $\frac{1}{3}$ for every —CH_2— group intervening between it and the point of attack. Figure 6 shows that the general relationship between *S* and the oxidation rate is maintained in the primary alkyl chlorides.

TABLE 3. THE INFLUENCE OF STRUCTURE ON THE RATE OF OXIDATION OF SOME CHLORO-PARAFFINS

structure	oxidation rate (relative to *n*-pentane)	stability factor
(*a*) primary chloro-paraffins		
$CH_3.\overset{*}{C}H_2.CH_2Cl$	0·2606	1·33
 $CH_3.\overset{\overset{CH_3}{\vert}}{\overset{*}{C}H}.CH_2Cl$	0·059	2·33
$CH_3.CH_2.\overset{*}{C}H_2.CH_2Cl$	3·62	0·66
 $CH_3.CH_2.\overset{\overset{CH_3}{\vert}}{CH}.CH_2Cl$	1·48	1·44
 $CH_3.\overset{\overset{CH_3}{\vert}}{CH}.\overset{*}{C}H_2.CH_2Cl$	2·41	1·0
$CH_3.CH_2.\overset{*}{C}H_2.CH_2.CH_2Cl$	6·45	0·44
(*b*) secondary chloro-paraffins		
$CH_3.CH_2.\overset{*}{C}HCl.CH_3$	1·77	0·79
$CH_3.CH_2.\overset{*}{C}H_2.CHCl.CH_3$	3·19	0·48

* Indicates assumed point of attack.

The introduction of a chlorine atom into the end of a hydrocarbon molecule may exert two distinct influences: (1) it impairs the symmetry of a methyl group, (2) it exerts its own specific inductive effect. These two effects are difficult to separate. The experimental results indicate clearly that a —CH_2Cl group as a whole exerts a less unfavourable influence than a —CH_3 group on ease of oxidation though the relative contributions of (1) and (2) cannot yet be estimated.

Secondary chloro-paraffins. The influence of the chlorine atom alone may, however, be found from measurements on secondary chloro-paraffins. The experimental results are unfortunately not numerous since, of the compounds examined, only 2-chloro-butane and 2-chloro-pentane undergo normal oxidation, the reaction of

2-chloro-propane with oxygen being probably an oxygen-catalyzed pyrolysis (see above).

Table 3 (*b*) shows that both the chloro-compounds studied oxidize about three times as fast as the corresponding hydrocarbons. Thus the replacement of hydrogen by chlorine facilitates oxidation, even when the substitution does not destroy the symmetry of a methyl group. In 2-chloro-butane: $CH_3.\overset{(1)}{CH_2}.\overset{(2)}{CHCl}.CH_3$, the favourable influence of chlorine makes attack more probable at (2) than at (1). If we ascribe to the chlorine atom a contribution y (relative to $CH_3 = 1{\cdot}0$), the stability factor of this compound will be: $0{\cdot}33+1{\cdot}0+y$. At the same time, it appears from figure 6 that for this compound (for which $\log_{10}(\text{rate}) = 0{\cdot}249$) the value for S should be 0·79, whence

$$0{\cdot}33+1{\cdot}0+y = 0{\cdot}79,$$

so that $y = -0{\cdot}54$.

In 2-chloro-pentane, $CH_3.CH_2.\overset{(1)}{CH_2}.\overset{(2)}{CHCl}.CH_3$, the most likely points of attack will be (1) and (2). If we adopt the above value of $-0{\cdot}54$ for the influence of chlorine, the stability factor will be a minimum for attack at (1). S is then $0{\cdot}33+0{\cdot}33+\frac{1}{3}(-0{\cdot}54)$, i.e. 0·48, which makes this compound conform to the general series (figure 6).

Assuming that the inductive influence of chlorine is the same in both primary and secondary chloro-paraffins, an estimate can now be made of the relative magnitudes of the two effects produced by substituting chlorine for hydrogen in a methyl group. At an adjacent carbon atom, the contribution of 1·0 for $—CH_3$ becomes 0·33 when this group is replaced by $—CH_2Cl$, so that the introduction of Cl causes a change of $-0{\cdot}67$ in the stability factor. Of this, $\frac{1}{3}(-0{\cdot}54)$, i.e. $-0{\cdot}18$ is due to the inductive influence of chlorine, so that the remainder, i.e. $-0{\cdot}49$ might be ascribed to the impairment of the symmetry of the methyl group. The latter effect thus appears to be of considerably greater importance.

Peroxide stability. Observations on the influence of structure show clearly that in the oxidation of hydrocarbons and alkyl chlorides (1) methyl groups have an unfavourable effect on the rate, and (2) substitution of chlorine for hydrogen facilitates oxidation, both because the number of methyl groups may thereby be diminished and because the chlorine exerts a specific influence of its own.

In an earlier paper (Cullis *et al.* 1947) the large structural influences encountered in hydrocarbon oxidation were ascribed primarily to the varying ease of fission of the O—O bond in intermediate peroxides. The stability of such peroxides may well depend on electronic effects transmitted by the groups attached to the two oxygen atoms. Chlorine, which is a powerfully electronegative element, constitutes an electron-attracting group. Its specific influence is opposite to that of the methyl group which possesses a capacity for electron-repulsion.

The results so far obtained therefore suggest that the O—O bond in peroxides of the type: $R—O—O—R'$ is strengthened if the attached groups are electron-repelling, and weakened if they are electron-attracting. This conclusion is in agreement with the view recently expressed by Walsh (1947, 1948), who argues that the strength of bonds between strongly electronegative elements will be increased by the attachment of electron-repelling groups, since such groups will facilitate expansion of the atomic orbitals and thereby allow increased overlap of the latter without the occurrence of nuclear repulsion.

REFERENCES

Coulson, C. A. 1947 *Quart. Rev. Chem. Soc.* **1**, 144.

Cullis, C. F., Hinshelwood, C. N., Mulcahy, M. F. R. & Partington, R. G. 1947 Faraday Soc. Discussion 'The labile molecule', p. 110.

Derick, C. G. 1911 *J. Amer. Chem. Soc.* **33**, 1181.

Hinshelwood, C. N. 1948 *J. Chem. Soc.* p. 531.

Rice, F. O. & Rice, K. K. 1935 *The aliphatic free radicals* (Chap. VII). Baltimore: Johns Hopkins.

Smith, J. O. (Jr) & Taylor, H. S. 1939 *J. Chem. Phys.* **7**, 390.

Steiner, H. & Watson, H. R. 1947 Faraday Soc. Discussion 'The labile molecule', p. 87.

Walsh, A. D. 1947 *J. Chem. Phys.* **15**, 688.

Walsh, A. D. 1948 *J. Chem. Soc.* pp. 331, 398.

A dynamical model of a crystal structure. II

BY SIR LAWRENCE BRAGG, F.R.S. AND W. M. LOMER
Cavendish Laboratory, University of Cambridge

(*Received* 27 *August* 1948)

[Plates 2 and 3]

A quantitative study of the mechanical properties of two-dimensional bubble rafts has been made. The elastic properties have been used to check the theory of the interaction of the bubbles, and reasonable agreement is obtained. The plastic properties are the main subject of this paper. Plastic deformation proceeds by the motion of dislocations, and we have taken advantage of the possibility of making absolutely perfect crystalline rafts to study the production of dislocations. They are produced either in pairs in the body of the rafts, or singly from the edges. The shear strain required to produce them varies from 7° for bubbles of 2 mm. diameter to a fraction of a degree for bubbles of less than 0·5 mm. diameter. The analogy with metals is best for bubbles of about 1·2 mm. diameter; for these the critical shear strain is about 3°. The observed shear strains can be explained in terms of the forces between the bubbles, and so it seems that the very low shear strengths of metallic single crystals cannot be explained in terms of dislocation creation, but must be a consequence of the initial presence of faults of some kind.

1. INTRODUCTION

In a recent paper entitled 'A dynamical model of a crystal structure' (Bragg & Nye 1947) an account was given of the behaviour of a model of a metal consisting of an assemblage of bubbles. The bubbles are between 2 and 0·2 mm. in diameter, and are formed on the surface of a soap solution. They are attracted together by capillary forces, and kept apart by the pressure at the interfaces where they touch. They crystallize under the influence of these forces as close-packed hexagonal sheets in two dimensions, and as structures of cubic or hexagonal closest packing in three dimensions. A mass of the bubbles exhibits many phenomena characteristic of metals, such as grain growth after being broken up into a mass of small crystals, and slip by the movement of dislocations when strained beyond the elastic limit. These phenomena were described in a qualitative way.

In the present paper, we attempt a quantitative treatment of the mechanical properties of a two-dimensional sheet of bubbles floating on the liquid surface. The variation with distance of the attractive force between two small bubbles has already been calculated by M. M. Nicolson (1949). The variation of repulsive force with distance at the point of contact of two bubbles has been calculated by one of us (W.M.L.). The expressions for these forces make it possible to calculate the deformations of a raft subjected to stresses, and so to obtain values for Young's modulus and Poisson's ratio. The calculated values of these quantities are compared with the values determined by experiments with an extended regular raft of bubbles, and it is found that agreement is satisfactory, indicating that the theory is a good approximation.

We go on to calculate the limiting strain beyond which we would expect the bubble raft to cease to react elastically and suffer permanent deformation due to slip, and to compare the theoretical limit with that which is observed experimentally. This comparison is inevitably much less precise than the comparison between the theoretical and experimental values of the elastic constants. The experimental determinations of the strain at which slip starts vary considerably, as they depend upon the conditions of experiment. Slip may start either by the creation of two dislocations in the interior of a perfect crystal which run in opposite directions, or by a dislocation starting from a weak point at an edge of the raft. We have calculated upper and lower limits for the limiting strain, on certain assumptions which lead to an upper limit about three times as high as the lower limit, and it is found that the observed points for bubbles of a range of diameters lie between these upper and lower extremes. The agreement is satisfactory to this extent. It shows that the mechanical properties of the two-dimensional raft of bubbles are in accord with what we would expect from the forces determined by the surface tension of the soap solution.

Details of the calculations are given by one of us (W.M.L.) in an accompanying paper, and only the results will be quoted in the present paper, which deals with the comparison between theory and experiment.

2. FORCES BETWEEN TWO BUBBLES

The attractive force between two bubbles of equal diameter has been calculated by M. M. Nicolson. In the calculations a simple expression for the curvature of the free liquid surface is used which is equivalent to assuming that the square of its slope is small compared with unity; this condition is satisfied to a sufficient approximation for bubbles which do not exceed 2 mm. in diameter. Nicolson's formula for the potential energy V due to the attractive force may be expressed in the form

$$V = Ta_0^2\psi(R/a_0)\,K_0(d/a_0), \tag{1}$$

where T = surface tension,

a_0^2 = Laplace's capillary constant $T/\rho g$,

R = radius of free bubble,

d = separation of bubble centres,

K_0 = Bessel function of zero order,

ψ = function evaluated numerically by Nicolson.

A measurement of the surface tension of the soap solution used in the experiment gave a value of 23 dynes/cm. With this value

$$a_0 = 1{\cdot}48\,\text{mm.}, \quad Ta_0^2 = 0{\cdot}0503\,\text{erg.}$$

Graphs of the functions K_0 and of $Ta_0^2\psi$ are given in the accompanying paper, and enable V to be calculated for any values of R and d.

In his original calculation of the repulsive force Nicolson assumed that two bubbles in contact retain a spherical form except at the flat interface where they touch. Experimental measurements of the modulus of elasticity of a sheet of bubbles gave values only one-quarter of those calculated on Nicolson's assumption, and a further examination of the conditions for equilibrium revealed the source of the error. The bubbles are distorted from the spherical form when they are pressed together by the attractive forces. The bubbles yield where they touch and expand elsewhere, and the elasticity modulus is less than it would be if this alteration in form were precluded. It is this yielding which enables the inner surface of the bubble in contact with the liquid to merge tangentially into the flat double-sided film at the interface. The same condition holds at the edge of the thin double-walled cap at

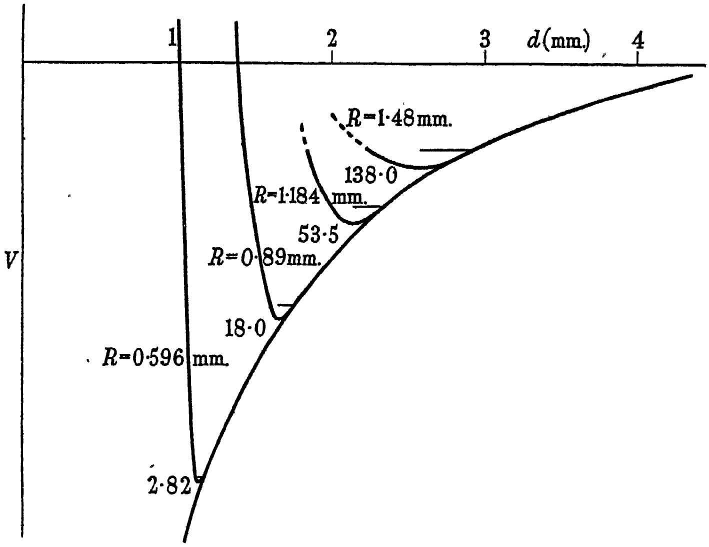


FIGURE 3. Abscissae represent the distance between bubble centres in millimetres. The ordinates are proportional to the mutual potential of pairs of bubbles, but for convenience in representing these potentials for bubbles of different sizes the actual quantities plotted are V/ψ (see equation (1)). The horizontal straight lines are drawn from the point of first contact to the point of equilibrium; the numbers at the minima give the mutual potential in ergs $\times 10^{-3}$.

the top of each bubble. Figure 1, plate 2, shows the fringes near the rim of this cap; the mean angle of the wedge as indicated by the first few fringes is less than $\frac{1}{4}°$. Another consequence of the deformation is that the repulsive force between two bubbles for a given value of d depends upon whether only a pair are in contact or whether both form part of an extended raft; it is greater in the latter case because expansion can only take place at top and bottom of each bubble. A recalculation of the elasticity allowing for this distortion gave good agreement with the observed value.

The repulsive force is due to the pressure inside the bubbles acting upon the interfaces. It is possible to see these interfaces through the film at the vertex of each bubble, which forms a cap or lid so thin that the interface is viewed without optical distortion. A photograph of interfaces is shown in figure 2, plate 2. The measured diameter of the circular rim of the interface is in good agreement with calculation, and this affords a check on the expression for the repulsive force, and on the calculated value of the attractive force which balances it.

Curves of potential energy for bubbles of different diameters are given in figure 3. It will be noticed that small bubbles are considerably more rigid than large ones.

3. THE ELASTIC PROPERTIES OF AN EXTENDED SHEET OF BUBBLES

A sheet of bubbles is elastically isotropic because of its hexagonal symmetry. Two elastic constants are therefore sufficient to specify its elastic properties. The most convenient constants to measure are Young's modulus and Poisson's ratio.

To measure Young's modulus, a raft of rectangular shape is formed between two smooth barriers on opposite sides, the other two sides being free. One barrier is attached to pivoted arms so that it is free to move; its small gravity-controlled deflexion is used to measure the stress after a calibration with a light rider on a special arm. The other barrier can be moved backwards or forwards, and its displacement, after a small correction for the yielding of the suspended side, measures the strain. Young's modulus is the ratio of the stress to the strain.

Poisson's ratio may readily be measured by taking two photographs, one of a free raft of bubbles and the other of the same raft which has been compressed or extended as in a Young's modulus measurement. If these photographs are superimposed as in figure 4, plate 2, a moiré effect is produced. If the raft is being compressed, spacings will be smaller in the direction of pressure, and greater at right angles to it. For some intermediate angle there will be neither compression nor extension, and this direction is shown by the stripes in the moiré pattern. If the slope which they make with the direction of compression or extension is θ, it can easily be shown that

$$\nu = (1+\cos 2\theta)/(1-\cos 2\theta).$$

Since Poisson's ratio ν cannot exceed unity, $\cos 2\theta$ is always negative. For instance, if Poisson's ratio were unity and the expansion of the raft sideways were equal to its compression in the direction of the applied force, the stripes would be at an angle θ of 45°. If Poisson's ratio were zero so that no sideways expansion took place, the

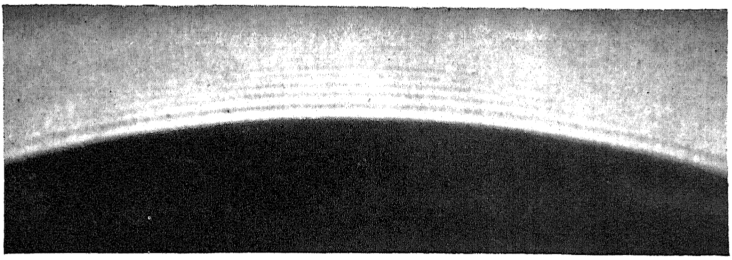


FIGURE 1

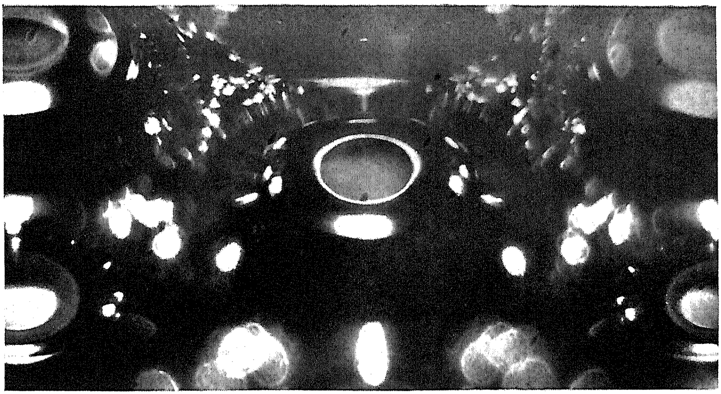


FIGURE 2

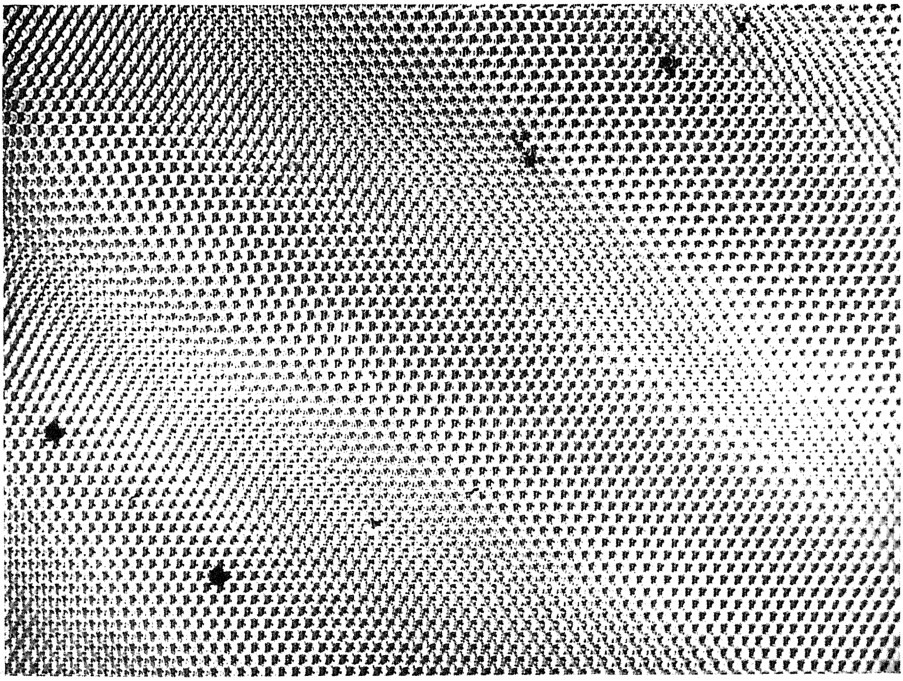


FIGURE 4

(*Facing p.* 174)

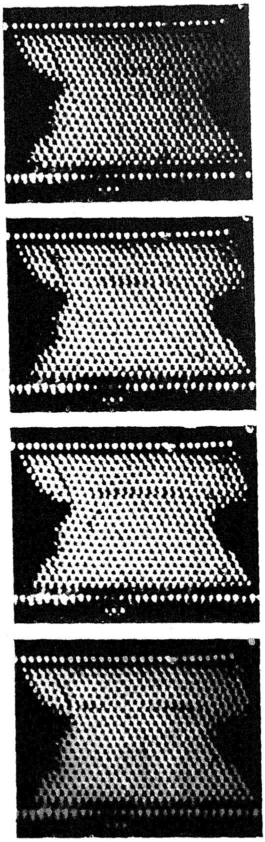
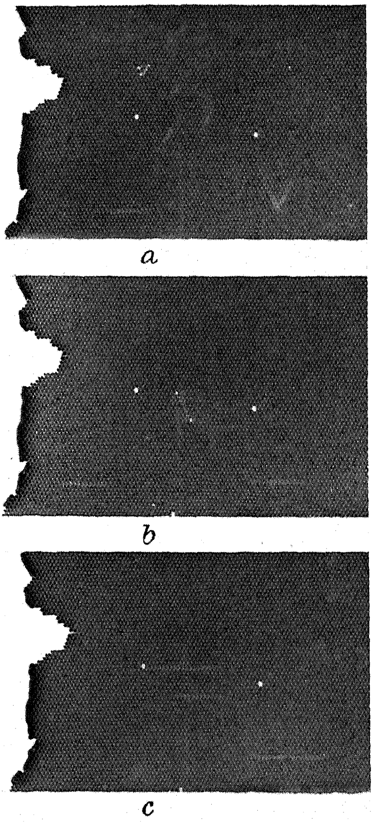


Figure 6



a

b

c

Figure 8

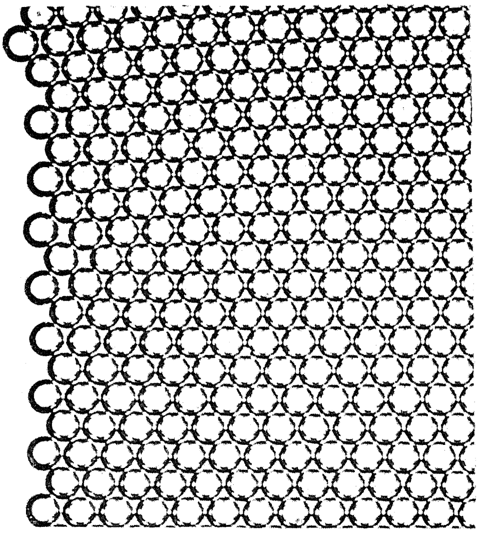


Figure 11

stripes would be perpendicular to the applied force ($\theta = 90°$). Actually they range around an angle of about 60°. The observed values of Poisson's ratio increase fairly uniformly with bubble diameter, from 0·45 to 0·6 when the diameter increases from 0·8 to 2·0 mm. The calculated values fit quite well up to a diameter of 1·5 mm.; after this the approximations of small strains and additive deformations break down and lack of agreement can be expected.

It may be noted that if there were independent central forces between the bubbles, Cauchy's relation would give $\nu = 0{\cdot}33$. The difference is due to interaction. A bubble which is pressed in at one point expands at other points of its surface as has been explained above.

The calculated and observed values of Young's modulus for different values of R are plotted in figure 5. The agreement is as good as might be expected, bearing in mind the approximations used in the calculations and the errors of measurement.

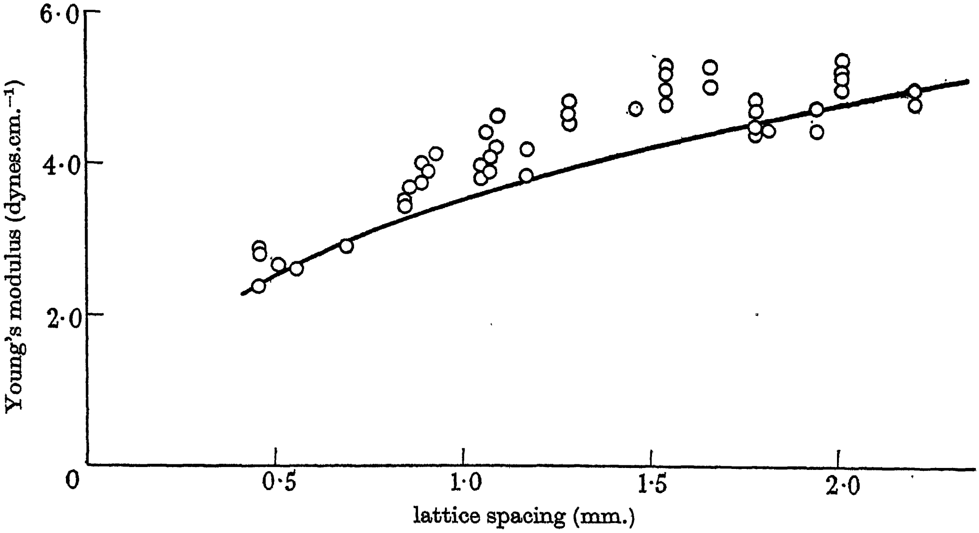


FIGURE 5. Young's modulus is plotted against bubble size. (The lattice spacing is approximately $2R$.) —— theory, ○ experimental results.

4. OBSERVATIONS ON THE ORIGIN OF DISLOCATIONS

The experiments on the formation of dislocations in perfect single crystals were carried out using one or other of two types of distortion, either simple uniaxial compression or tension between smooth barriers as in the Young's modulus experiment, or else shear applied by forming the raft between two parallel springs of pitch equal to the bubble diameter and translating one of them parallel to its length. In both cases we noted the range between the strains in either direction at which slip began.

The latter experiment might at first sight appear to be of the more simple type, but this is not the case. The stresses set up are complicated. For instance, if the free edges of the raft are perpendicular to the springs there can be no shear stress acting along the slip lines where they meet the edges. An edge is therefore not likely to give rise to single dislocations and deformation should proceed by the formation of pairs of dislocations in the body of the raft. The creation of such pairs was in fact observed. Figure 6, plate 3, shows successive frames from a 16 mm. cinematograph film taken at intervals of $\frac{1}{16}$ sec. Two dislocations are formed which then run in opposite directions to the free edges. The points of creation of the dislocations were in satisfactory qualitative agreement with the positions of the maxima of shear stress as deduced by Inglis (1923) and by Mann (to be published) (see figure 7).

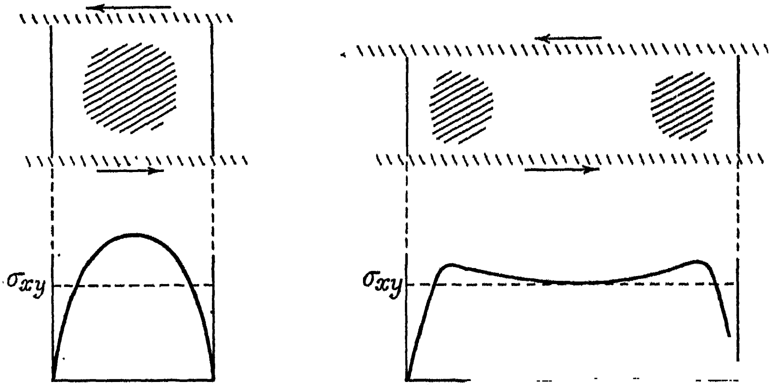


FIGURE 7. The curves give the values of the shear stress along a median line parallel to the springs; the broken line represents the shear stress which would be deduced simply from the relative movement of the springs. Most dislocations start in the shaded areas where the strain is greatest.

The experiment is not suited for quantitative comparison, because of uncertainty as to the exact fit of the rows of bubbles in the space between the springs and consequent uncertainty of the compression normal to the slip lines. It did, however, appear that slip started at a fairly definite shear strain whatever the dimensions of the raft. This observation contradicts a statement in the previous paper (Bragg & Nye 1947, p. 480), that slip appeared to start when energy was released by its completion. It would now appear that this condition was merely true for rafts of the dimensions used in the earlier experiments, and does not hold in general.

More consistent experimental results are obtained when a regular raft is subjected to simple uniaxial compression or extension between smooth barriers. The free edges of the raft were trimmed so as to be perpendicular to the barriers. It was compressed or extended until slip occurred along one of the lines inclined at 60° to the barriers. The overall length of the elastic range was measured for various sizes of bubbles. Results were reasonably reproducible provided care was taken to make the free edges straight. Dislocations always started from an edge, never in pairs in the body of a perfect raft. If a point of weakness, such as a small notch, were deliberately made at one edge it became a very probable origin of dislocations owing to the local stress concentration. Even in such a case, the critical strain was seldom lowered by a factor as great as two. Vacant lattice sites gave rise to pairs of dislocations in compression, becoming filled up in the process, but their presence had no marked effect on the critical strain. If a dislocation parallel to the barriers happened to be present, it turned on slight compression into two dislocations inclined at 60° and as compression proceeded these ran away along appropriate slip lines. A fixed boundary reflected a dislocation which ran into it so that a single dislocation could cross the whole raft by reflexion backwards and forwards between the barriers. A cinematograph film of these phenomena has been made.* In it a raft of small bubbles is compressed and extended through a range much greater than the elastic range. The dislocations appear as dark lines, because the displacement of the bubbles exposes the blackened bottom of the dish containing the solution. The dislocations can be seen darting about along the slip directions, starting at points of stress concentration at an edge, being reflected at the barriers, being created in pairs at faults in the structure, or running together and creating a fault, and in fact presenting an animated scene which gives a graphic picture of what may be supposed to be taking place in a strained metal crystal. A series of phenomena of this type is shown in figure 8, plate 3, which shows three frames taken from the cinematograph film. The raft was being extended in the vertical direction. The V in the lower right-hand corner of figure 8*a* becomes a dislocation in figure 8*b*, and this lengthens as the tension increases. Two parallel dislocations in the centre of figure 8*a* run together to form a complex fault in figure 8*b*, and this develops into two parallel dislocations on a slip plane inclined at 60° to the first. A dislocation is seen in the bottom left-hand corner which runs towards the edge. In these prints from a positive film the dislocations appear as white lines.

5. EXPERIMENTAL DETERMINATION OF THE ELASTIC LIMIT

The measurements of the elastic limit for uniaxial tension or compression are plotted in figure 9.

The range of strain between a compression at which slip starts and an extension at which slip starts is plotted as ordinate against the diameter of the bubbles as abscissa. Table 1 gives the mean shear strain along the slip lines deduced from the compressional and extensional strains. It must be remembered, however, that this cannot be simply correlated with the critical shear strain for an infinite lattice, because in this experiment the dislocations start from a free edge. It is interesting to note how rapidly the critical shear strain diminishes as the bubbles become smaller and behave as more rigid bodies.

TABLE 1

diameter (mm.)	2·0	1·5	1·2	1·0	0·6
critical shear strain (observed)	6·65°	4·7°	3·25°	2·7°	1·1°

We should have liked to have pursued the investigation with still smaller bubbles, but such small strains cannot be measured except with a large raft, and this contains such a very large number of the small bubbles that it is difficult to manipulate into perfect regularity before shrinking and bursting of the bubbles defeats the experiment.

---

* We are greatly indebted to Mr A. P. M. Brookes of the Department of Engineering, Cambridge University, on whose expert knowledge we have depended for the making of the cinematographs.

6. COMPARISON OF THEORETICAL AND EXPERIMENTAL VALUES OF LIMITING SHEAR STRAIN

An upper limit to the theoretical value to be expected can be calculated as follows. Let it be supposed that when slip takes place, all the bubbles in one row move simultaneously over those in the opposite row on the other side of the slip plane. The initial state of equilibrium of the two rows is represented by figure 10*a*, where the spheres are shown as intersecting to indicate the formation of the flat areas of contact of the bubbles which give rise to the repulsive forces.

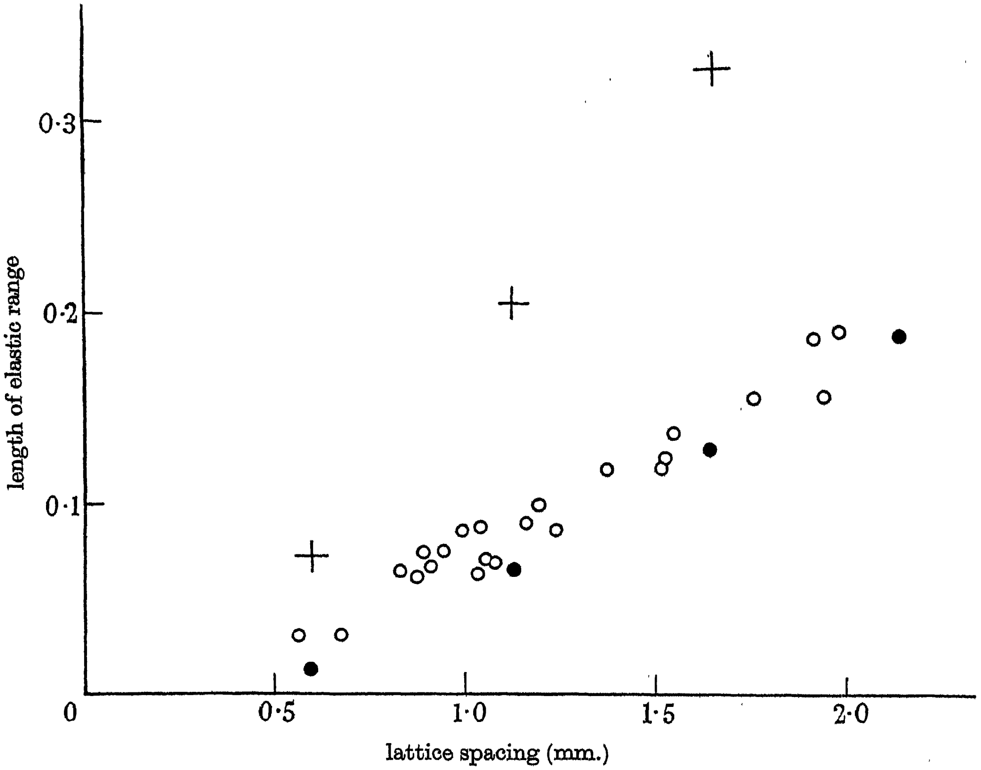


FIGURE 9. Comparison between experimental determination and theoretical prediction of the length of the elastic range, for a range of bubble diameters. ○ measured, + theoretical upper limit, ● theoretical lower limit.

In figure 10*b* the shearing movement has taken place to such an extent that each of the upper bubbles is just losing contact with its lower left-hand neighbour, so that the repulsive force has fallen to zero. There is then equilibrium between the shearing force which is drawing the upper row of bubbles to the right, the repulsive forces at the remaining points of contact, and the internal pressure arising from the attractive forces which is pressing the two rows together. Beyond this point, the upper bubbles may be thought of as sliding over the lower ones up an incline which becomes less steep as shear proceeds, so that the shearing force diminishes and finally falls to zero when one row is vertically above the other. The shearing force must have attained its maximum value at the strain corresponding to the point where contact is just being lost as in figure 10*b*. A crystal uniformly strained beyond this point is unstable; the slightest excess of shear on any one glide plane makes that plane weaker than its neighbours, so that slip takes place on it while shear strain is released on the neighbouring planes. The strain calculated in this way therefore represents an upper limit to the elastic range.

The shorter the range of the repulsive force, the lower will be the critical angle of shear. It is clearly for this reason that the critical angle diminishes as the bubbles become smaller and more rigid. In fact, as Orowan has pointed out in discussion, the critical shear strain should be zero for an assemblage of incompressible spheres. This argument is only given in a general way here, and is developed more fully in the accompanying paper.

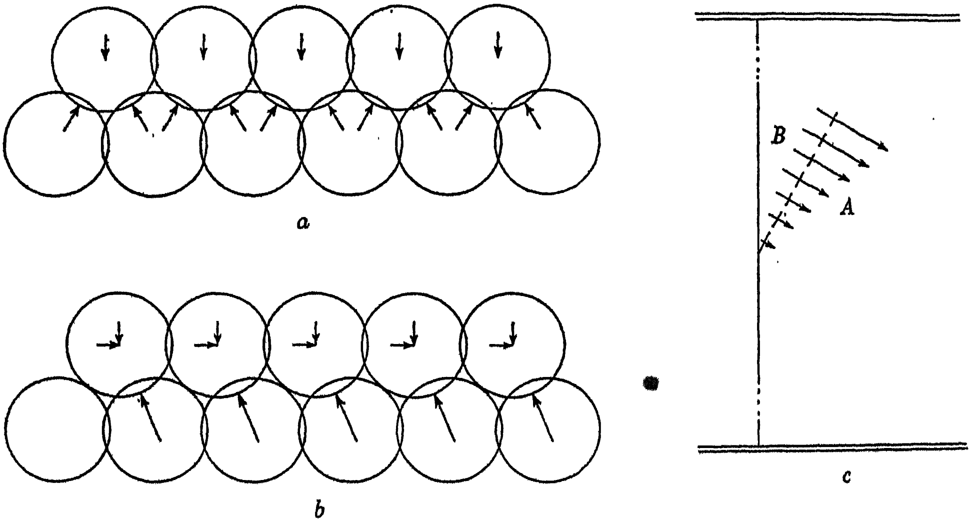


FIGURE 10

In the experiment on the critical values of compression and extension the slip started at an edge, and this has two distinct effects. The attractive forces which give rise to the internal pressure are long range forces as compared with those of repulsion, and hence the internal pressure diminishes as the edge of a raft is approached. The last rows of bubbles at the edge are held to the raft by a force which is only a fraction (e.g. 0·3 for 0·6 mm. bubbles) of the attraction across a slip line in the interior. When considering slip starting from one edge along the line of dashes in figure 10*c*, it is to be expected that it will be controlled by the smaller value of the attractive force between *A* and the bubbles at the tip of the wedge *B*. We have calculated a lower limit to the critical strain on this assumption. The second effect is only operative in compression. It is shown in the acompanying paper that at a certain value of the compression a bubble which is at the extreme edge of the raft will be squeezed out of position. The movement of a bubble at an edge is supposed to initiate a dislocation. This process has actually been observed with medium sized bubbles of 1·4 mm. diameter. An example of its occurrence is shown in figure 11, plate 3. A lower limit has been calculated for the range between the onset of slip in tension and compression, allowing for the diminution in the attractive force and, in the case of compression, assuming a dislocation to start by the displacement of a single bubble at the edge.

These upper and lower limits are plotted in figure 9. It will be seen that points representing the observations lie between the calculated curves in the case of the smaller bubbles, and approximate to the lower limit in the case of the larger ones.

7. ANALOGY WITH METALS

To sum up, the observed strength of a perfect bubble raft is in agreement with the 'theoretical strength', calculated in terms of the forces between the bubbles. The observed critical strains, which may be quite low for small bubbles, are not the consequence of pre-existent dislocations or special points of weakness. An irregular edge or vacant lattice site lowers the limit, but not to less than half its value for the perfect raft.

We undertook this investigation in an attempt to cast some light on the factors which determine the yield point in a metal. It is interesting to note that the critical yield strain may be quite low in the bubble model. We have observed values as low as 1° for bubbles of 0·5 mm. diameter; an extrapolation of our curve indicates that the critical angle for an infinite raft of bubbles of 0·15 mm. diameter would be about 13 min. of arc. This value is not far from the critical shear strain in copper which has been hardened by cold-work, which is about 10 min. of arc according to Taylor & Quinney (1934).

However, the comparison is not as promising as this correspondence might indicate. In the first place, our experiments have been made with a perfect raft. Comparison ought to be made with a single crystal of copper which yields for almost vanishingly small values of stress. In the second place, critical strains of the order of a degree or less are only observed with very small bubbles. It is shown in part III that the forces between copper atoms in copper metal correspond most closely to those between bubbles of 1·2 mm. diameter. This correspondence is arrived at by choosing a diameter for the bubbles which gives the closest correspondence between its law of repulsive force and the calculations of Fuchs on the variation of repulsive force with interatomic distance for copper ions. The critical shear strain for bubbles of 1·2 mm. diameter is 3·25°, a value some twenty times greater than the value even for hardened copper.

Since bubbles of quite large size best represent the metal atoms, and such bubbles have quite high values of critical strain if the crystal is essentially perfect, it would appear unlikely that stresses cause a metal to slip by creating dislocations in an initially perfect structure.

Measurements on the properties of rafts containing dislocations are not as profitable as might be thought. The bubbles have effectively no mass, and therefore a dislocation has no kinetic energy. The main method of energy dissipation is by work against viscous forces, whereas in a metal it is presumably by the radiation of elastic waves. Experiments to determine the minimum strain required to move a dislocation have not been successful. The viscosity determines the rate at which a dislocation proceeds. Smaller strains cause it to move at even slower rates, and the experiment is rendered impossible by the time limit imposed by the shrinking and bursting of the bubbles.

We therefore cannot consider the bubble model as representing the dynamic behaviour of a real lattice, quite apart from its disadvantage of being two- and not three-dimensional. We nevertheless consider that its behaviour is interesting, if only because it shows that calculations of the 'theoretical strength' lead in fact to values in good accord with observation.

REFERENCES

Bragg, W. L. & Nye, J. F. 1947 *Proc. Roy. Soc.* A, **190**, 474.
Inglis, C. E. 1923 *Proc. Roy. Soc.* A, **103**, 568.
Mann, E. M. 1949 *Proc. Camb. Phil. Soc.* (in the Press).
Nicolson, M. M. 1949 *Proc. Camb. Phil. Soc.* (in the Press).
Taylor, G. I. & Quinney, H. 1934 *Proc. Roy. Soc.* A, **143**, 307.

EXPLANATION OF PLATES 2 AND 3

PLATE 2

FIGURE 1. Fringes in the wedge of liquid near cap.

FIGURE 2. Flat interfaces seen through cap of bubble. (Illuminated from underneath to get specular reflexion from flats.)

FIGURE 4. Superposition of two plates of different compressions to show direction of zero compression.

PLATE 3

FIGURE 6. The formation of a pair of dislocations, from 16 mm. film.

FIGURE 8. Three frames from 16 mm. film of raft being extended.

FIGURE 11. Squeezing out of edge bubbles of a raft in compression—the first stage in the formation of a dislocation (illuminated from below).

A dynamical model of a crystal structure. III

By W. M. Lomer, *Cavendish Laboratory, University of Cambridge*

(*Communicated by Sir Lawrence Bragg, F.R.S.—Received* 27 *August* 1948)

A calculation of the maximum shear strain under which a two-dimensional close-packed lattice is stable has been carried out in terms of the forces between the lattice components. Two types of force were used; those between floating bubbles, which enabled a comparison with experiments on actual rafts of bubbles to be made, and also the forces derived from a potential $V = A\,e^{-\beta r^2}$, which form has been frequently proposed as an approximation to the repulsive interaction terms between metal ions. The conclusion reached is that this maximum strain may be considerably less than that deduced from a simple sine law approximation to the shear force versus displacement curve. Detailed consideration is given to edge effects in bubble rafts, and reasonable agreement with experimental results obtained. The overall result is that the formation of dislocations and consequent plastic yielding can occur in an initially perfect lattice only at quite large shear strains. The analogy with metals is discussed, and we conclude that the low strengths of metallic single crystals are explicable only on the assumption that they are not perfect and that dislocations already exist in them and move under very small shear stresses.

1. Introduction

In this paper the shear strain at which a perfect two-dimensional lattice of bubbles becomes unstable is calculated. The calculation is applied to actual experimental conditions; the comparison of the results of theory and experiment are given in the accompanying paper. The forces between pairs of bubbles on which the whole calculation is based are very briefly described; a fuller account of their derivation and confirmation will be published elsewhere. A discussion of the analogy with metals shows that medium sized bubbles should best represent a typical metal such as copper. There are many rather questionable assumptions and approximations in the calculation, but the main purpose is to show that in a perfect lattice the strength is of the same order as the theoretical value despite the fact that plastic deformation proceeds by the motion of dislocations, and that at the low elastic strains obtainable in metals there is little possibility of the stresses producing dislocations.

2. Forces between bubbles

These forces have been discussed in § 2 of part II. The attractive force acting between a pair of bubbles is derived from a potential (see equation (1) of part II)

$$V = Ta_0^2\psi\left(\frac{R}{a_0}\right)K_0\left(\frac{d}{a_0}\right) \equiv \Phi K_0\left(\frac{d}{a_0}\right), \tag{1}$$

where d = separation of bubble centres, a_0^2 Laplace's capillary constant. Φ is a function of the size of the bubbles and of the solution used, which we evaluate numerically from Nicolson's work (1949), while K_0 is the zero order Bessel function. Both Φ and K_0 are plotted in figure 1. In the present calculation we do not use the attractive forces directly in this form, but instead represent their effect in an infinite raft as an internal pressure whose value is found by summing many terms of the form (1). This internal pressure gives rise to a force of F dynes per bubble; F decreases rapidly as the bubble diameter decreases. Since many comparatively distant bubbles contribute to F we assume it to be independent of local deformations.

The repulsive forces, on the other hand, are determined solely by the positions of nearest neighbours, the magnitude of the force between any pair being the product of the excess pressure $2T/R$ and the area of the flat film between them. Because a bubble compressed in one direction expands in other directions the force between two bubbles is not a function only of their separation but also of the positions of their other neighbours. It may be assumed that, if γ is the ratio of the expansion at 60° to the direct compression, the force f_i along a close-packed direction in a hexagonal array will be

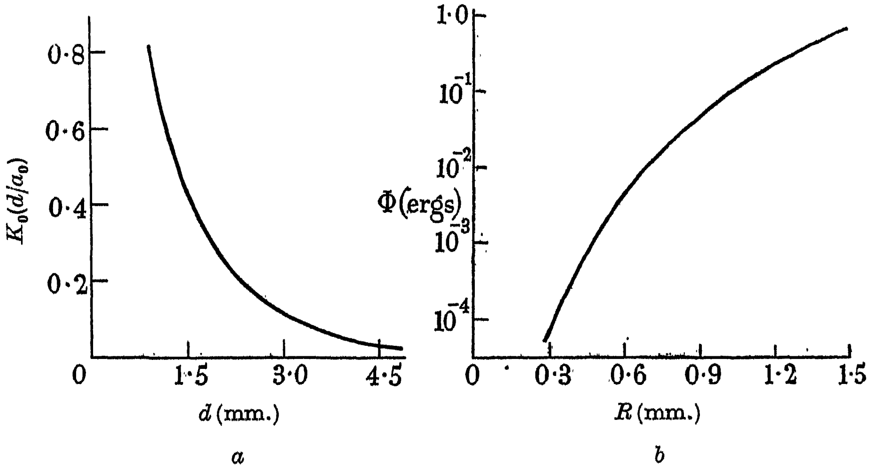


FIGURE 1. a The function $K_0(d/a_0)$ is plotted against the separation of bubble centres d ($a_0 = 1{\cdot}48$ mm. for the solution used). b The function Φ (equation 1) is plotted logarithmically against the bubble radius.

$$f_i = f(\Delta_i) = f(\delta_i + \gamma\{\delta_j + \delta_k\}), \tag{2}$$

where i, j, k denote the three close-packed directions, and Δ_i is the 'overlap' of two bubbles lying in the ith close-packed direction. This overlap can be regarded as made up of two contributions; first, the approach of the centres of the two bubbles concerned makes an overlap equal to the difference between the free diameter $2R$ and the actual separation d, and secondly it will be increased by expansions in this ith direction arising from compressions in the other two directions j and k. The first contribution, expressed as a strain, is written $\delta_i = (2R-d)/2R$; the compressions in the other two directions are δ_j and δ_k and so the second contribution is the sum of the expansions $\gamma\delta_j$ and $\gamma\delta_k$. Thus, if the deformations are additive, which is probably true only for small total deformations, the overlap Δ_i is $\delta_i + \gamma\{\delta_j + \delta_k\}$ and is greater than the apparent value $(2R-d)/2R = \delta_i$.

The function $f(\Delta)$ has been calculated and is plotted in figure 2. Using this curve we can find the value of f_i corresponding to any set of δ_i, δ_j, δ_k.

The ordinates of figure 2 are dimensionless and so apply to all sized bubbles; they must be multiplied by $\frac{1}{2}\pi TR$ to get the actual repulsive force in dynes.

When the raft is in equilibrium (denoted by suffix 0), the repulsive forces f balance the attractive force F, and we easily see (figure 4b) that

$$f_0 = \frac{F}{\surd 3} = f(\Delta_0). \tag{3}$$

Since in equilibrium all three axes i, j, k are equivalent we have

$$\Delta_0 = \delta_0(1+2\gamma). \tag{4}$$

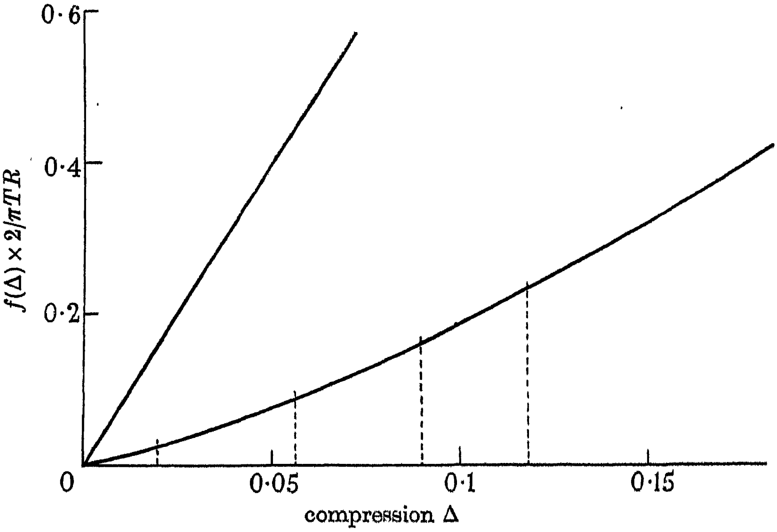


FIGURE 2. The ordinates are a dimensionless representation of the force resulting from a uni-axial compression plotted as abscissa. The vertical broken lines show the equilibrium compressions or overlaps Δ_0 in an infinite raft for bubbles of radii (from left to right) of 0·296, 0·592, 0·888 and 1·184 mm.

This quantity δ_0, the fractional difference between the lattice spacing in an infinite raft and the diameter $2R$ of a free bubble, proves to be very important; it is the key factor in the calculation of the shear strength and most other properties of bubble rafts depend on it in some degree. δ_0 is bigger for larger bubbles, as is shown by the vertical lines in figure 2 and by the second column of table 1. When δ_0 is small, the bubbles are nearly spherical; when it is large, they are more nearly hexagonal in cross-section. It is the close approximation of the smaller bubbles to spheres that gives the impression of great rigidity and hardness.

3. SHEAR STRENGTH OF A PERFECT INFINITE LATTICE

The experiments described in part II showed that the shear strain at which plastic yielding of perfect bubble rafts begins depends very much on the size of the bubbles, ranging from about 1° to 10° for the sizes used. In much theoretical work (e.g. Peierls 1940; Nabarro 1947) the variation of force as a row of atoms is displaced parallel to the adjacent row is assumed to be approximately sinusoidal, the value of the shear modulus of elasticity fixing the scale of the curve. On such a supposition, we find that the shear strain at which the shearing force begins to decrease with increasing strain, that is, the strain for which the crystal becomes unstable, corresponds to a relative displacement of adjacent rows ϕ equal to a quarter of the lattice constant a. This is a shear strain of $\frac{1}{4}a.2/\sqrt{3}.a$ and is 16°. Thus either the sine law is a poor approximation, or else other effects, such as those arising from the edges, are active to various degrees for different sized bubbles.

It was found possible to measure the force required to shear two whole rows over one another by using a modification of the parallel spring arrangement described in § 4 of the accompanying part II. Only two rows of bubbles were used and one spring was suspended so that its deflexion measured the force. The relative movement of the rows was deduced from the motion of the micrometer screw with a small correction for the deflexion of the suspended side. A comparison between the experimental figures for bubbles 1·3 mm. diameter and the sine curve which best fits the elastic region is made in figure 3.

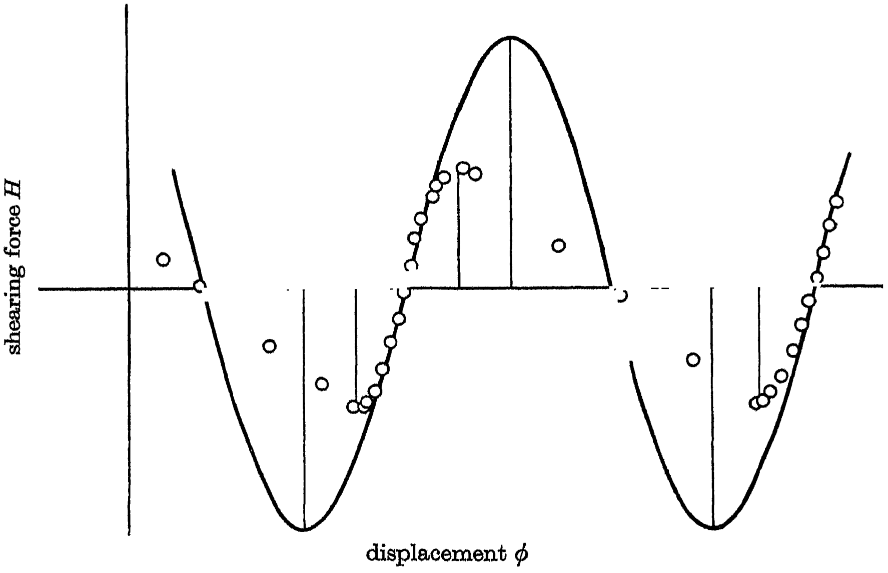


FIGURE 3. The shear force acting when two adjacent rows are displaced parallel to one another. —— sine curve fitted to elastic region; ○ experimental points.

It is seen that the maximum force, which denotes the onset of instability, occurs at a much smaller displacement than the $\frac{1}{4}a$ of the sine approximation. An approximate calculation of the position of this maximum is made in terms of the forces outlined above.

Figure 4a shows the forces acting on a bubble in a raft. In the elastic region $f_1 = f_4; f_2 = f_5; f_3 = f_6$; when the raft is free from stress the symmetry shows all these forces to be equal to f_0. Taking a section parallel to a row along the broken line we find a resultant force normal to the row equal to $\sqrt{3}.f_0$. This is balanced by the attractive forces whose resultant we have called F (figure 4b). There are no tangential forces since the horizontal components of f_5 and f_6 are equal and opposite.

Now consider the raft under shear strain. The forces still balance in pairs, but the forces f_2 and f_3 are no longer equal, and a tangential force H acts on the section because of the difference in their horizontal components. This force represents a shear stress and is transmitted by the raft to the edges where it is balanced by the applied forces. The force F still balances the vertical components of f_2 and f_3.

We want to estimate the shear strain at which the force H reaches a maximum. A full solution of the problem would be extremely complicated. Suppose we give the raft a shear strain which displaces the bubble B a distance $\frac{1}{2}a-\phi$ parallel to a row as shown in figure 4*c*. The equilibrium position of the bubble will, in general, be such that its displacement is not purely parallel to the row; the angles of the lattice will have changed, and so too will the forces f_2 and f_3. A simplified treatment which enables us to find the displacement for the maximum shear stress fairly accurately is given.

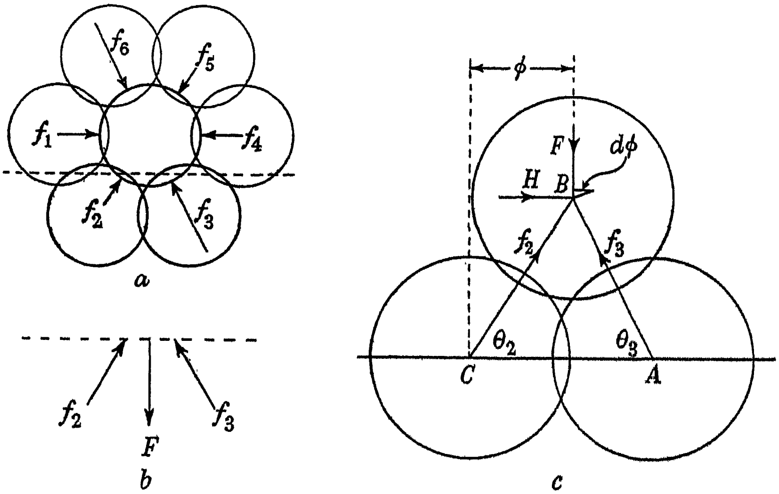


FIGURE 4

Consider the equilibrium of the bubble B. The resultant of the forces f_1, f_6, f_5 and f_4 has components F normal, and H parallel, to the rows. These are balanced by the forces f_2 and f_3. The force F, as was explained in § 2, is considered to be independent of the distortion, while H is the shear force whose maximum we investigate. Take moments about the centre of A. This gives, with $BC = a$,

$$Fa\cos\theta_3 = Ha\sin\theta_3 + f_2 a\sin(\pi-\theta_2-\theta_3).$$

Suppose the strain of the raft is increased, so that the change in equilibrium position of B along the row is $d\phi$. The changes in the moments give

$$-F\,d\phi = d(Ha\sin\theta_3) + d(f_2 a\sin\{\theta_2+\theta_3\}),$$

or $$-F\,d\phi - H d(a\sin\theta_3) - f_2 d(a\sin(\theta_2+\theta_3)) = a\sin\theta_3\,dH + a\sin(\theta_2+\theta_3)\,df_2.$$

Now if H is at its maximum, $dH = 0$. Also df_2 can be written $(df_2/d\Delta)\,d\Delta$; we know $df/d\Delta$, the change in force due to a certain compressive strain, from the curve of figure 2, and $d\phi$ gives a strain $d\Delta$ along the join BC of $-\cos\theta_2(d\phi/a)$. (This should be slightly increased because of the movement of the bubble normal to the row. It can be seen that this would make df_2 greater and so our argument even better.) Since θ_2 is nearly 60° this is $-d\phi/2a$.

So $$-F\,d\phi - H\,d(a\sin\theta_3) - f_2\,d(a\sin(\theta_2+\theta_3)) = -a\sin(\theta_2+\theta_3)\frac{df}{d\Delta}\frac{d\phi}{2a}.$$

The third term is very small, and the second is about $\frac{1}{2}H\,d\phi$ and H cannot be more than $F/\sqrt{3}$, so that certainly within a factor of two the equation can be written

$$2F = \frac{df}{d\Delta}.$$

From figure 2 we see that, on the arbitrary scale there used, F is never greater than 0·3. So the slope should not be greater than 0·6; at $\Delta = 0{\cdot}02$ the slope is 1·3, so that even in the most unfavourable circumstances the maximum of H occurs when the compression between two bubbles B and C is about 1 %, which is less than a tenth of the equilibrium compression, and we can therefore use the criterion that the force f_2 shall vanish to define the state of maximum shearing stress. The value of the force H at this point is then

$$H = f_3\cos\theta_3. \tag{5}$$

Also we have $$F = f_3\sin\theta_3, \tag{6}$$

therefore $$H = F\cot\theta_3. \tag{7}$$

This relation holds as long as $f_2 = 0$, and so applies to all the region where two rows ride over one another, giving the form of the falling part of the stress-strain curve. It is almost linear.

The lattice as a whole is unstable when it is at the point of maximum shear stress, for if one plane gets an accidental extra displacement the stress on it will decrease and it will continue to move, relieving the strain in the rest of the raft. We now see the importance of the quantity δ_0; it is this which determines the strain necessary just to separate bubbles B and C.

To proceed further it is necessary to make other approximations which limit the calculations to small bubbles. The measured strains are small for small bubbles, and this justifies treating the angles of the lattice in the critical state of strain as being nearly 60°. Subject to this approximation we have from equation (5) and equation (6),

$$H = F/\sqrt{3}, \quad f_3 = \frac{2F}{\sqrt{3}}.$$

So the critical state is defined by the system of forces

$$f_1 = f_0, \quad f_2 = 0, \quad f_3 = 2f_0.$$

We must find the compressions δ corresponding to these using equation (2) and curve 2, and use these to give the shear strain

$$e_{xy} = (\delta_2 - \delta_3)\frac{2}{\sqrt{3}} = \frac{2}{\sqrt{3}}\left(\frac{1}{2} - \frac{\phi}{a}\right). \tag{8}$$

It is apparent from figure 2 that the force curve cannot be considered linear over ranges of compression as large as δ_0. We call δ_{III} the compression, in direction 3, which gives $f_3 = 2f_0, f_2 = 0$. This is calculated from the actual force curve figure 2 and involves some interaction terms. The sides of a triangle such as ABC in figure 4c are then

$$2R(1-\delta_0), \quad 2R(1-\delta_{\mathrm{III}}), \quad 2R(1+\gamma(\delta_0+\delta_{\mathrm{III}})),$$

and an application of the cosine formula gives

$$\frac{1}{2}-\frac{\phi}{a}=\frac{1}{2}-\frac{\phi}{2R(1-\delta_0)}=\frac{\delta_{\mathrm{III}}(1+\gamma)+\gamma\delta_0}{(1-\delta_0)^2}. \tag{9}$$

In table 1 values of δ_0, δ_{III} and e_{xy} for several values of bubble diameter are given and it is seen that the maximum theoretical shear strain is about two or three times δ_0, the homogeneous expansion required to make the repulsive forces vanish. Thus for the very hard small bubbles we get low critical strains.

TABLE 1

R (mm.)	δ_0	δ_{III}	e_{xy}	e_{xx} tension	e_{xx} compression	squeezing out e_{xx}	$\exp\left(-\frac{\sqrt{3}R}{a_0}\right)$	e_{xx} lower limit	n
0·296	0·0185	0·035	0·048	0·0268	0·0477	0·005	0·7	0·013	12·5
0·592	0·044	0·086	0·134	0·0815	0·132	0·023	0·5	0·064	5·0
0·888	0·076	0·136	0·220	0·129	0·196	0·045	0·35	0·129	3·2
1·184	0·103	0·170	0·270	0·162	0·245	0·065	0·25	0·186	2·8

4. MODIFICATIONS FOR ACTUAL CONDITIONS OF EXPERIMENTS

(i) *Change to tensile tests*

It is not yet possible to make a comparison with the experiments since these were not done on infinite rafts in pure shear, but for finite rafts under uniaxial tension and compression, such that slip occurred along lines at 30° to the compression direction. The application of the stress not only provided a shear component, but also altered the value of F, the force acting normal to the slip line. If F_0 be the force due to the attractive forces, we can write

$$F = F_0 + 2R\sigma \sin^2 30,$$

where σ = stress = force per unit length, and so requires the factor $2R$ to reduce it to the force per bubble. Thus

$$F = F_0 + \tfrac{1}{4}\sigma . 2R.$$

The tangential force H is

$$H = \sigma \sin 30 \cos 30 . 2R = \frac{\sqrt{3}}{4}\sigma 2R.$$

The maximum value of H is, as above, $F/\sqrt{3}$, so that for slip to occur

$$\pm\frac{\sqrt{3}}{4}\sigma\,2R = \frac{1}{\sqrt{3}}\left(F_0+\frac{\sigma}{4}\,2R\right),$$

that is

$$2R\sigma = -F_0 \quad \text{or} \quad 2F_0. \tag{10}$$

The negative sign indicates tension.

Thus in tension $F = \frac{3}{4}F_0$, and in compression $F = \frac{3}{2}F_0$.

The value of F is introduced into the calculation of the critical strain via δ_0, the initial compression. Thus by using different values of δ_0 and δ_{III} we can obtain values for the strain from the above equations (8) and (9); this can be converted to linear extension by the formula

$$e_{xx} = \frac{2}{\sqrt{3}}\frac{1}{1+\nu}e_{xy}, \tag{11}$$

and directly compared with experiment.

There are, however, two objections to this. First, a tension of the value indicated is sufficient to pull the raft apart normally to the tension. This was observed, but only for the smallest bubbles. Secondly, the strength in compression was not observed to be greater than that in tension, which is the implication of the different values of F above. The former objection is not serious, for we have neglected the variation of the angles θ (figure 4*c*) by taking $H = F/\sqrt{3}$. This would reduce the critical value of H by several per cent if it were taken into account especially for larger bubbles where the θ undergo larger variations. Thus the brittle fracture of small bubbles and its absence for larger ones is quite satisfactory. The second objection is more serious, but we believe that *as a calculation of the strength of an infinite raft* the above is valid. One effect of the edges, as we shall show, is to make a greater reduction in the compressive strength than in the extensional. The results, using the above values of $F = \frac{3}{4}F_0$ and $F = \frac{3}{2}F_0$ are given in the fifth and sixth columns of table 1; their sum is plotted in figure 9 of part II as the theoretical upper limit. They are necessarily an upper limit, for we have neglected edge effects and considered the angles θ constant.

Before proceeding to the edge effects, we must discuss the relevance of these calculations of the displacement of whole rows to the formation of dislocations. It is clear that if a crystal is uniformly at the maximum stable shear strain a small additional local strain will make the crystal yield in that region; the stress in this region will decrease, and so more will be supported by the surrounding parts of the lattice and the plastic region spreads. An alternative way of stating this is that the extra strain lifts over a part of a row into its next equilibrium position, thus producing a pair of dislocations of opposite signs which separate under the applied stress. It is obvious that such pair creation involves a temporary local increase in elastic energy, and so we suppose that pair formation will occur only if there is a local weakness or concentration of shear stress. This is in accord with the observations described in part II; that pairs were not formed in uniform compression, and that in the shear experiments, they formed at peaks of shear stress. The production of a single dislocation at the edge of a raft is easily understood if it is assumed that the theoretical strength is rather less there than elsewhere. Then two rows get displaced relative to one another at the edge and this displacement again spreads, the edge of the displaced region being a dislocation. In the following section it is assumed that the main difficulty is the formation of dislocations, and not their movement, and accordingly we take the strength of the lattice to be determined by that strain which exceeds anywhere in the lattice the theoretical strain calculated on the above lines.

(ii) *Edge effects*

The edge has two distinct effects. One is again represented as a variation of the attractive force F; this force is less near the edge since, as explained in part II, there are fewer bubbles beyond the slip line to be attracted to the raft as a whole. The reduction in F due to this effect at the edge is such that we get

$$F = F_0\left(1-\exp\left(-\sqrt{3}\,\frac{R}{a_0}\right)\right). \tag{12}$$

The application of this equation is difficult, for the distance into the raft in which the force F reaches its maximum value is of the same order as the length of a dislocation. We ought therefore to use an intermediate value between F and F_0, but it is by no means easy to decide what mean to take; the value of $\exp(-\sqrt{3}\,R/a_0)$ is given in column eight of table 1 and shows the order of magnitude of the effect.

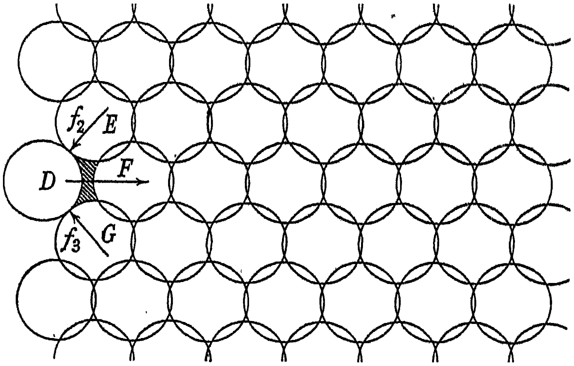


Figure 5. The squeezing out of an edge bubble in a compression experiment (see also figure 11 of part II).

The second effect of the edge is active only in compression, and is illustrated in figure 5. The bubble D is situated at the edge, and when the raft is compressed, the forces f_2 and f_3 increase as E and G approach one another. Now when the horizontal components of f_2 and f_3 together exceed the attractive force $F_0\left(1-\exp\left(-\sqrt{3}\,\frac{R}{a_0}\right)\right)$ the bubble will move outwards as shown in figure 5 until the forces f_2 and f_3 have decreased to give equilibrium again. This decrease in f_2 will upset the equilibrium of E to a certain extent, and it may be that E will start to follow D, thus initiating a dislocation. This mechanism has actually been observed for medium sized bubbles ~1·4 mm. diameter (see for instance part II, figure 11). The strain to displace D has been calculated (column seven, table 1), and this is considered to be a lower limit to the compressive strength, because the movement of bubbles in a position like that of D must be the first stage in the formation of any dislocation. This strain, added to that for tension (using the reduced edge value of F), gives the lower limit to the strength (column nine, table 1 and part II, figure 9).

For the smallest bubbles, to which the assumptions most nearly apply, the experimental points lie between our upper and lower limits, while for the larger ones, the lower limit seems almost correct. This is reasonable since for the larger bubbles dislocations are only some two or three bubbles long, so that the displacement of the first one, our lower limit, is in effect equivalent to the formation of a dislocation. Also, the basic calculation assumed the angles of the lattice to be constant, an assumption which, as we noted earlier, will tend to make all of the calculated strengths higher than they should be. Thus it appears that the agreement between theory and practice is quite as good as could be expected.

5. LENGTH OF DISLOCATIONS

Another rather crude check of the theory can be made by considering the length of dislocations. Let us define the length as the number of spaces where corresponding bubbles have lost contact, i.e. where the relative displacement of the two adjacent rows is greater than the critical value. As noted by Bragg & Nye (1947) this length increases rapidly as the bubble diameter decreases. The experimental data are very poor, for the length depends almost entirely on boundary conditions; in figure 6 the arrows show the sort of spread of lengths even when as much care as possible is taken in the experimental conditions. The theoretical curve is derived as follows.

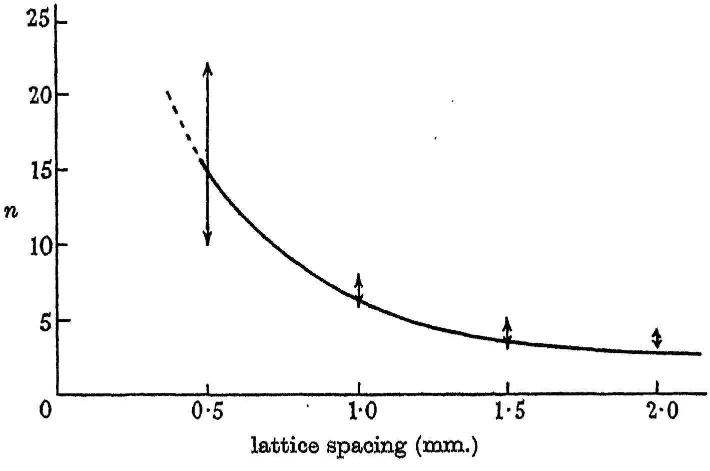


FIGURE 6. The length of dislocations is plotted against lattice spacing. —— theory; arrows observations.

Koehler (1941) has shown that the shear strain at a distance r from a dislocation (r being in the direction of the displacement vector of the dislocation) is

$$e_{xy} = \frac{ma}{2\pi(m-1)}\frac{1}{r},$$

where $1/m = \nu =$ Poisson's ratio.

This solution is valid as long as Hooke's law is true; in our case we assume this is so until the critical strain is reached. Using the measured values of Poisson's ratio we can thus find the value of r for which the critical shear strain is exceeded. Then $2r/a$ is the number n of spaces visible, that is

$$n = \frac{m}{\pi(m-1)} \frac{1}{e_{xy}}. \tag{13}$$

The curve in figure 6 was plotted using the calculated values of e_{xy} (see last column of table 1); there is reasonable agreement.

To sum up, then, it may be claimed that the yield strength of an initially perfect bubble raft can be explained in terms of the known forces which are acting between the bubbles. It is not very much less than the theoretical strength of a lattice, that is, the shear required to make two whole rows displace relative to one another simultaneously. The shear strains required to form dislocations are about 5° for the sizes of bubbles used, and so are far above the yield strains of metals. We could get smaller strains by using smaller bubbles, and so must consider now what sizes best represent metals.

6. ANALOGY WITH METALS

The behaviour of the bubbles is largely determined by their size, and we now try to find what size best represents a typical metal. We compare only the repulsive forces, since the attractive ones may be represented in both cases as an internal pressure.

Some calculations by Fuchs (1936) enable a direct comparison with copper. He calculates the potential arising from the closed-shell interactions of the copper ions, and this is compared with the potential due to the repulsive forces between two bubbles. The ratio of slope to ordinate of the curves at the equilibrium spacing is fitted to Fuchs's curve as well as possible; it is found that this needs an equilibrium compression Δ_0 equal to 0·055 for the bubbles. For a diameter of 1·2 mm. $\Delta_0 = 0{\cdot}0565$, and so this size is used for calculation of the potential curve, which is of course the integral curve of figure 2. The general form is quite a good fit for Fuchs's points, but this is in part deceptive, and the curvatures at the equilibrium distance which determine the elasticity do not agree very well. However, this quantity does not enter into our calculation and so it seems that as far as the formation of dislocations is concerned the elasticity will not be important. For these bubbles we obtained a critical shear strain $= 0{\cdot}134$ radians.

Now Fuchs's figures have been approximated by a formula $V = A\,e^{-\beta r^2}$. Using such a form we can follow Nabarro (1947) and get an expression for the energy of mutually displaced rows

$$V = \frac{B}{\beta}\left[1 + \log\frac{A\beta}{B} + \log\sum_{n=-\infty}^{\infty} e^{-\beta(na+\phi)^2}\right].$$

B is a measure of the normal force we have called F. ϕ is the only variable in this equation, and so a direct differentiation gives us the tangential force $H = -dV/d\phi$.

By considering only nearest neighbours we reduce the sum to two terms, and finally find that the maximum of H occurs when

$$\frac{\phi}{a} = \frac{1}{2} \pm \frac{1}{2\beta a^2}\cosh^{-1}(\beta a^2 - 1).$$

The shear strain is given by $$e_{xy} = \frac{2}{\sqrt{3}}\left(\frac{\phi}{a} - \frac{1}{2}\right).$$

As $\beta \to \infty$, $e_{xy} \to 0$ so that again we see that infinitely hard lattice components have zero critical shear *strain*. For copper $\beta a^2 \sim 20$ and so $e_{xy} = 0{\cdot}11$.

This is smaller than the bubble value, probably because no interaction was considered in the metal case. The value for the bubbles is reduced to 0·112 when γ is made zero. Thus it seems that the size bubble to represent copper ions is about 1·2 mm. diameter; the length of the dislocations is then about five atoms, which agrees with many previous estimates, although detailed comparison is not easy owing to the ill-defined term 'length'.

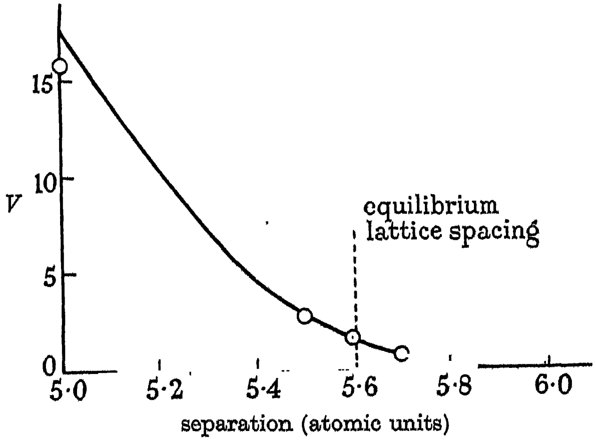


FIGURE 7. The potential due to repulsive forces is plotted against the separation. ○ Fuchs's calculations for copper ions; —— potential curve for bubbles of diameter 1·18 mm. reduced to same scale as Fuchs's results.

The potential curves for alkali metals and inert gases could not be fitted by any size of bubble, the slope at the equilibrium distance being a smaller multiple of the ordinate than for any bubbles. We can use the second argument, however; for the alkalis $\beta a^2 \sim 7$ and so e_{xy} is about 0·21. This is reached (without interaction) by bubbles about 2·0 mm. diameter and so this size might be that to use. If this is so, dislocations will be about three atoms long and the elastic range of a single crystal almost double that for copper.

The conclusion to be drawn from this work seems to be fairly definite. Bubbles of quite large sizes best represent metal ions; they have very high yield strains if the crystal is initially perfect. It is therefore unlikely that stress alone can produce dislocations, and since the thermal fluctuation energy kT is some hundred times smaller than the estimated self-energy of a dislocation, even the joint effect of stress and fluctuations seems a doubtful source of dislocations in a perfect lattice.

Measurements on the properties of rafts containing dislocations are not as informative as might be hoped. There is no doubt that a dislocation runs along its slip line under very small shear stresses. The measurement of the minimum stress to set a dislocation in motion was not possible; in the bubble model, the strain energy which the dislocation relieves is dissipated in work against viscous forces, and the slower it moves the less work it dissipates. In practice the shear strain can be reduced and the dislocation made to travel more slowly until it is moving so slowly as to appear stationary—some would move again after a stop over a minute. The time limit of bursting and shrinking again puts a stop to this sort of experiment. This need cause no great disappointment, since in any case the model cannot give a good representation of the dynamics of a metal; the lattice components have no mass and no thermal motion, dislocations have no kinetic energy, and energy dissipation against viscous forces replaces radiation of elastic waves. These considerations, however, should not invalidate the conclusions we have drawn about the formation of dislocations being unlikely, and even the limitation to a two-dimensional system should not have any serious qualitative effects.

REFERENCES

Bragg, W. L. & Nye, J. F. 1947 *Proc. Roy. Soc.* A, **190**, 474.
Fuchs, K. 1936 *Proc. Roy. Soc.* A, **153**, 622.
also Mott & Jones 1936 *Properties of metals and alloys*, p. 143. Oxford University Press.
Koehler, J. S. 1941 *Phys. Rev.* **60**, 397.
Nabarro, F. R. N. 1947 *Proc. Phys. Soc.* **59**, 269.
Nicolson, M. M. 1949 *Proc. Camb. Phil. Soc.* (in the Press).
Peierls, R. 1940 *Proc. Phys. Soc.* **52**, 34.

Vibration spectra of hydrocarbon molecules

I. Frequencies due to deformation vibrations of hydrogen atoms attached to a double bond

BY N. SHEPPARD AND G. B. B. M. SUTHERLAND

Laboratory of Colloid Science, University of Cambridge

(*Communicated by Sir John Lennard-Jones, F.R.S.—Received* 18 *May* 1948)

An analysis has been made of the infra-red and Raman spectra of alkyl substituted ethylenes for 'characteristic' frequencies in the region between 1500 and 700 cm.$^{-1}$. The mean values of the frequencies characterizing the various types of substitution are as follows: asymmetrically disubstituted, 1415 and 890 cm.$^{-1}$; *trans*-disubstituted, 1303 and 973 cm.$^{-1}$; *cis*-disubstituted, 1260 and 973 cm.$^{-1}$; mono-substituted, 1415, 1295, 990 and 910 cm.$^{-1}$; trisubstituted, 1383 and 820 cm.$^{-1}$. These frequencies have all been assigned to specific modes of vibration, essentially localized in deformation motions of the hydrogen atoms directly attached to the double bond; the higher frequencies (>1000 cm.$^{-1}$) are concerned with motions in the plane of the ethylenic double bond and the lower frequencies arise from motions out of that plane. Using a valency force field, a theoretical explanation has been given of the persistence of these characteristic frequencies in alkyl substituted ethylenes and detailed calculations have been made of the relevant force constants involved. The variation in certain of these characteristic frequencies when halogen atoms replace the alkyl groups has been considered and shown to be due to changes in the force constants and not to the changes in the masses of the substituents. For a given type of substitution, the force constant decreases progressively along the series alkyl group, iodine, bromine, chlorine and fluorine, i.e. with increasing electronegativity. The use of these methods to provide a means of following changes in the electronic structure of the double bond with substitution is discussed.

INTRODUCTION

The fundamental vibration frequencies of hydrocarbon molecules fall naturally into five classes. Proceeding from high to lower frequencies these are:

(1) a group of frequencies between 3400 and 2850 cm.$^{-1}$ (2·9 to 3·5μ);

(2) a narrow band of frequencies near 2000 cm.$^{-1}$ (5μ);

(3) a narrow band of frequencies near 1650 cm.$^{-1}$ (6μ);

(4) a moderately wide band of frequencies between 1500 and 600 cm.$^{-1}$ (6·7 to 16·7μ);

(5) a group of frequencies between 450 and 200 cm.$^{-1}$ (22 to 50μ).

The general physical interpretation of these subdivisions has long been clear; thus, the first three groups of frequencies are due to vibrations which are highly localized respectively in the stretching of C—H, C≡C and C═C bonds. The fourth group, however, includes two entirely different types of motion, viz. (1) vibrations due to deformation motions of the hydrogen atoms in C—H bonds, and (2) vibrations due essentially to the stretching of C—C bonds. The fifth group arises from deformation modes of vibration of carbon atoms. Although the classification may break down in a few special cases (e.g. where resonance between neighbouring multiple bonds occurs or in certain ring structures) it certainly holds for the great majority of hydrocarbons and forms the natural starting point for any discussion of the vibration spectra of hydrocarbons as a whole. It is clear, however, that this general interpretation of the spectra does not take us very far in the correlation of the spectrum of an individual hydrocarbon with the structure of that hydrocarbon and it is only when detailed assignments of frequencies within these groups have been achieved that vibration spectra can be used to give reliable information on the structure and properties of hydrocarbons.

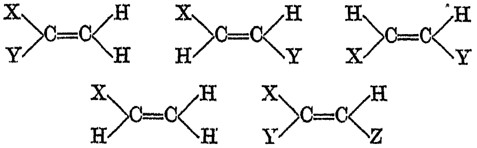
There are two lines of approach to this problem of assignment. The first is to concentrate on the simplest hydrocarbons, obtain a detailed analysis of the normal vibrations, investigate the selection rules governing transitions in infra-red and Raman spectra and so achieve a full interpretation of the spectra of a number of simple hydrocarbons. Having done this, one may hope to extrapolate the results to the more complex hydrocarbons for which the normal vibration analysis is impracticable. While this approach has given a theoretical justification for the above general classification of frequencies, the extrapolation of the results on simple hydrocarbons to more complex molecules has so far led to little of importance. The main reason is that each molecule studied has usually had special characteristics of its own, particularly some form of symmetry, not existing in the larger molecules. A further reason is that the internal force fields controlling the vibrations have not yet been established with certainty, so that even in the simpler hydrocarbons controversy still exists over some of the assignments.

The second, and at present more promising, line of approach is to isolate the vibrations of certain key groups or unit structures common to classes of hydrocarbons and to concentrate attention on the detailed interpretation of the normal vibrations of these groups. The justification for this method is that in practice it has been found that the spectra of hydrocarbons possessing certain key groups do in fact have many spectroscopic features in common. It seems reasonable to conclude that the internal vibrations of these key groups are, to a high degree of approximation, independent of the rest of the molecule and may therefore be isolated for separate theoretical treatment. In particular, certain vibrations of the hydrogen atoms are virtually independent of the vibrations of the carbon atoms, and Fox & Martin (1940) have shown how frequencies in the first class (CH stretching) may be dealt with in this manner to characterize CH, CH_2 and CH_3 groups in particular environments in a wide variety of hydrocarbons. In the present two papers, we propose to consider the detailed theoretical interpretation of frequencies in class (4) in certain key groups existing in hydrocarbons. This first paper deals essentially with deformation vibrations of hydrogen atoms directly attached to carbon atoms in an ethylenic type of double bond; the second will be mainly concerned with C—C stretching vibrations in paraffinic hydrocarbons containing terminal groups of the type $CH_3.CH(CH_3).CH_2...$ or $CH_3.C(CH_3)_2.CH_2....$ In each case a large accumulation of experimental data had shown that frequencies characterizing these groups persisted over a wide range of hydrocarbon molecules containing them.

A preliminary note giving some of the results of the second investigation has already appeared (Simpson & Sutherland 1947).

THE EXPERIMENTAL DATA

The first requirement for our investigation is a set of reliable infra-red and Raman spectra of several hydrocarbons of each of the following structural types:



where X, Y and Z denote alkyl groups.

We do not propose to consider here the spectrum of the parent ethylene molecule as this has recently been fully dealt with by Herzberg (1945) with whose interpretation we are in essential agreement. Moreover, the relevance of the ethylene spectrum to those of the substituted ethylenes is not as great as might be expected.

The necessary data on the Raman spectra have been available for several years in the work of Bourguel and his associates (1931, 1932*a*, 1932*b*, 1933*a*, 1933*b*, 1935) and of Grèdy (1935*a*, 1935*b*, 1935*c*, 1937) who jointly made a detailed investigation of such hydrocarbons with the object of following the variation of the C=C stretching frequency with substitution, but these spectra are equally useful for our purpose. A certain number of these hydrocarbons have been re-examined by Cleveland (1943*a*, 1943*b*, 1943*c*, 1944) and more recently an extensive compilation of Raman spectra of hydrocarbons has been published (Fenske and others 1947). We have used these more detailed results where applicable.

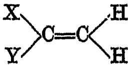
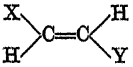
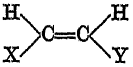
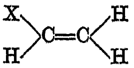
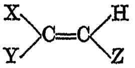
The infra-red spectra of series of hydrocarbons of these types were first obtained by Lambert & Lecomte (1932, 1935, 1938*a*, 1938*b*) and Andant, Lambert & Lecomte (1935), but during the past six or seven years the whole field has been covered with much higher resolving power by several workers in this country (Sutherland & Thompson 1945; Fellgett, Harris, Simpson, Sutherland, Thompson, Whiffen & Willis 1946) and in U.S.A. (Rasmussen & Brattain 1947 and *American Petroleum Institute Reports* 1947).

Certain of the absorption bands between 1000 and 800 cm.$^{-1}$ in these infra-red spectra were quickly correlated with characteristic olefinic groupings and have been widely used since as an empirical guide to molecular structure. The earliest work on these correlations was by Lambert & Lecomte (1938*a*, *b*) but Rasmussen & Brattain (1947) and Thompson & Torkington (1945) have derived a more complete set of rules from the newer data. It is generally recognized that these very important infra-red bands are associated with deformation modes of the hydrogen atoms but no detailed assignments have been made. One of the main purposes of this paper is to show how these and other such bands can be assigned to specific modes of vibration.

The infra-red and Raman spectra of a representative number of each of the various types of substituted ethylenes are shown in figures 1 to 9. The various sources from which these spectra have been taken have been enumerated above. In table 1 are given the characteristic frequencies that can be picked out from the spectra as common to each of the various types of substituted ethylenes. These frequencies can be divided into two series, (*a*) those that lie in the region 1250 to 1450 cm.$^{-1}$, most of which are strong in Raman scattering but medium or weak in intensity in infra-red absorption, and (*b*) those that lie between 800 and 1000 cm.$^{-1}$, all of which are strong in absorption and weak in scattering. We shall show that each of these groups of frequencies represents related vibrations of the molecules concerned.

However, before proceeding to the detailed discussion of these frequencies, it will be convenient to note two frequencies which occur in most of the spectra but which are characteristic of the saturated hydrocarbon chains of the molecules. These lie in the regions 1360 to 1380 and 1440 to 1480 cm.$^{-1}$ respectively, and are caused by deformation vibrations of hydrogens in the CH_2 and CH_3 groups. The 1360 to 1380 cm.$^{-1}$ frequency is well known to be the symmetrical deformation frequency of the methyl group, whilst bands in the 1440 to 1480 cm.$^{-1}$ region are caused both by the doubly degenerate deformation mode of this group, and by the symmetrical deformation of CH_2 groups (Herzberg 1945).

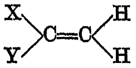
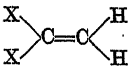
TABLE 1. OBSERVED CHARACTERISTIC FREQUENCIES IN THE VIBRATION SPECTRA OF OLEFINS

structural type	observed characteristic frequencies (cm.$^{-1}$)	intensity in infra-red (IR) and Raman (R)
	1410–1420	R, strong
	885– 895	IR, strong; R, weak
	1295–1310	R, strong*
	1295–1310	IR, medium*
	965– 980	IR, strong; R, weak
	1250–1270	R, strong
	965– 980	IR, strong; R, weak
	1410–1420	R, medium
	1290–1300	R, strong
	985– 995	IR, strong; R, weak
	905– 915	IR, strong; R, weak
	1380–1390	R, strong
	800– 840	IR, strong

* The reason for the assignment of these two frequencies to separate vibrations will be given later.

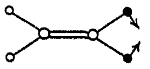
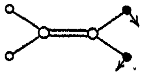
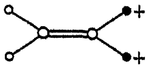
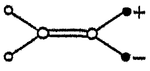
ASSIGNMENT OF THE CHARACTERISTIC FREQUENCIES

Several considerations lead us to assign these frequencies to hydrogenic vibrations of the structural units concerned. In the first place, the fact that they are independent of the precise nature of the alkyl groups X, Y, Z makes it appear unlikely that any of them are stretching vibrations of C—C linkages around the double bond. In the second place, it is generally apparent that as we proceed from mono-substituted to multiply-substituted ethylenes, the number of frequencies decreases with the number of hydrogen atoms remaining attached to the double bond. Furthermore this is true, whether we consider the whole range of frequencies or the two sub-classifications previously mentioned. The reverse effect would be expected if these frequencies arose from C—C stretching vibrations. We may best proceed to assign them to particular deformation modes of the hydrogen atoms by considering first certain of the symmetrical disubstituted ethylenes.

Molecules of the type  *and* 

For such molecules there will be four hydrogenic deformation frequencies associated with the double bond. The approximate forms of vibrations together with the selection rules for the more symmetrical molecules with two X groups (point group C_{2v}) are shown in table 2. In this and all subsequent tables, the hydrogen atoms are indicated by ● and heavier atoms (or groups of atoms) by ○. The approximate motions of the hydrogen atoms only are indicated.

TABLE 2. THE ASSIGNMENT OF THE CHARACTERISTIC FREQUENCIES OF ASYMMETRICALLY DISUBSTITUTED ETHYLENES

designation	form of vibration	symmetry class	selection rules	assignment	intense in
ν_1		A_1	IR and R	1410–1420	Raman
ν_2		B_2	IR and R	?	—
ν_3		B_1	IR and R	885–895	infra-red
ν_4		A_2	R	?	—

We shall consider in turn the assignment of the two observed frequencies near 1415 and 890 cm.$^{-1}$ (figures 1 and 2) to these four possible modes. The strong Raman frequency between 1410 and 1420 cm.$^{-1}$ can be assigned with confidence to ν_1 for two reasons. It is the strongest Raman line of those that we are considering, and hence would normally be assigned to the only totally symmetric vibration, viz. ν_1. Secondly, the corresponding symmetrical deformation of the CH_2 group in ethylene and in saturated hydrocarbons is well established at about 1450 cm.$^{-1}$.

The strong infra-red band (and weak Raman line) in the region 885 to 895 cm.$^{-1}$ is less easy to assign to one of the remaining three modes. However, ν_4 can be rejected as it would be inactive in the infra-red in the case of molecules of this type, whereas in practice the 890 cm.$^{-1}$ band remains strong in such molecules, e.g. isobutylene. We are thus left with a choice between ν_2 and ν_3 and in this case it proves to be possible to distinguish between these alternatives by considering the rotational contour of this infra-red band of isobutylene. An approximate calculation of the moments of inertia of isobutylene shows that it is very nearly a symmetrical top having almost equal moments of inertia about axes in the plane of the molecule. In the notation of Gerhard & Dennison (1933), $I_C \doteqdot 2I_A$ and hence $\beta = \frac{I_A}{I_C} - 1 \doteqdot -0{\cdot}5$.

Thus 'perpendicular' bands will have approximate contours consisting of two *P* and *R* shoulders with a weak *Q* branch in the middle, whereas 'parallel' bands will have very pronounced *Q* branches whose intensity will be nearly half that of the whole band.

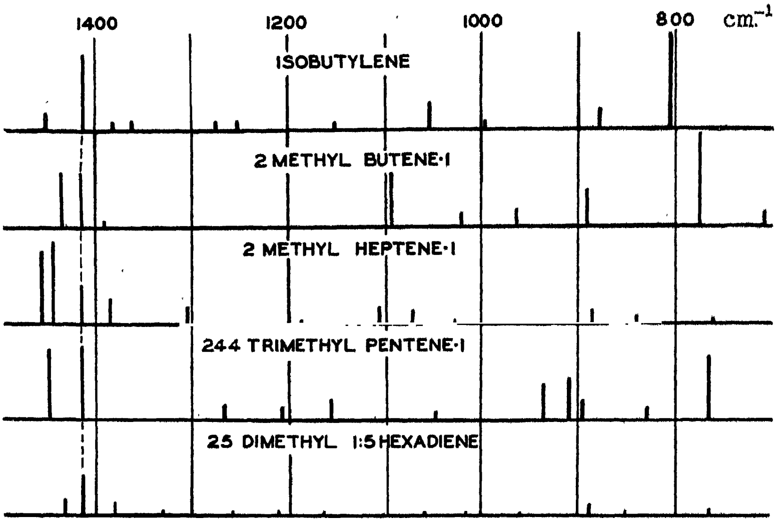


FIGURE 1. Raman spectra of asymmetrically disubstituted ethylenes.

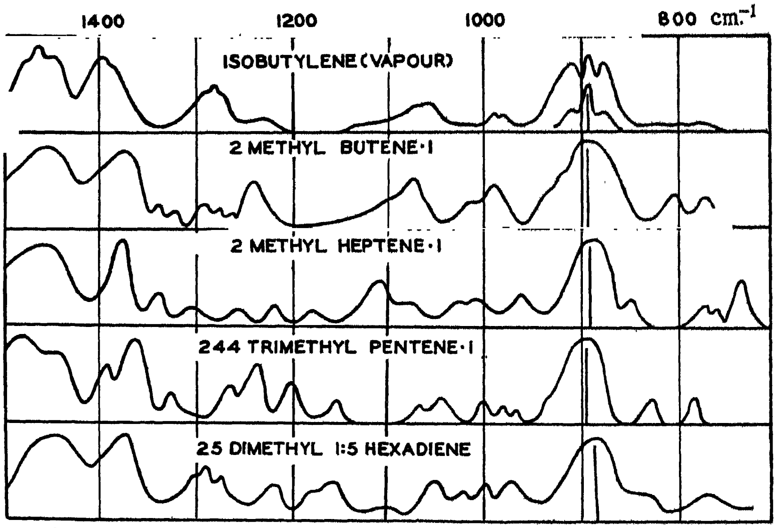
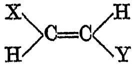
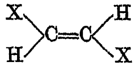


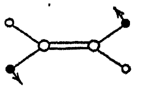
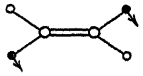
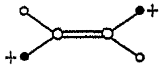
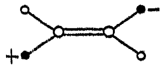
FIGURE 2. Infra-red spectra of asymmetrically disubstituted ethylenes.

Examination of the contour of the infra-red band at 890 cm.$^{-1}$ in isobutylene reveals that this has the predicted contour of a parallel type band. Hence the corresponding vibration has its associated change in electric movement perpendicular to the plane of the molecule and it must therefore be ν_3. It may be added that the calculated value for the *P*–*R* spacing of such a parallel band is approximately 35 cm.$^{-1}$ agreeing well with the observed spacing of 33 cm.$^{-1}$. As further confirmation of the above assignment of this absorption band we see that the corresponding infra-red active vibration in ethylene occurs at 949 cm.$^{-1}$ and is very intense in the infra-red, whereas the ν_2 analogue in ethylene is at a higher frequency (995 cm.$^{-1}$) and is much weaker in absorption. The remaining two modes ν_2 and ν_4 cannot yet be assigned with certainty to observed frequencies.

Molecules of the type  *and* 

The experimental data are given in figures 3 and 4 and the modes of vibration in table 3. The Raman data for *cis*- and *trans*-butene-2 are taken from Gershinowitz & Wilson (1938). Here again there are four possible modes to correlate with the observed frequencies at 1295 to 1310 cm.$^{-1}$ and 965 to 980 cm.$^{-1}$.

TABLE 3. THE ASSIGNMENT OF THE CHARACTERISTIC FREQUENCIES OF *TRANS* DISUBSTITUTED ETHYLENES

designation	form of vibration	symmetry class	selection rules	assignment	intense in
ν_1		A_g	R	1295–1310	R
ν_2		B_u	IR	1295–1310	IR
ν_3		A_u	IR	965–980	IR
ν_4		B_g	R	?	—

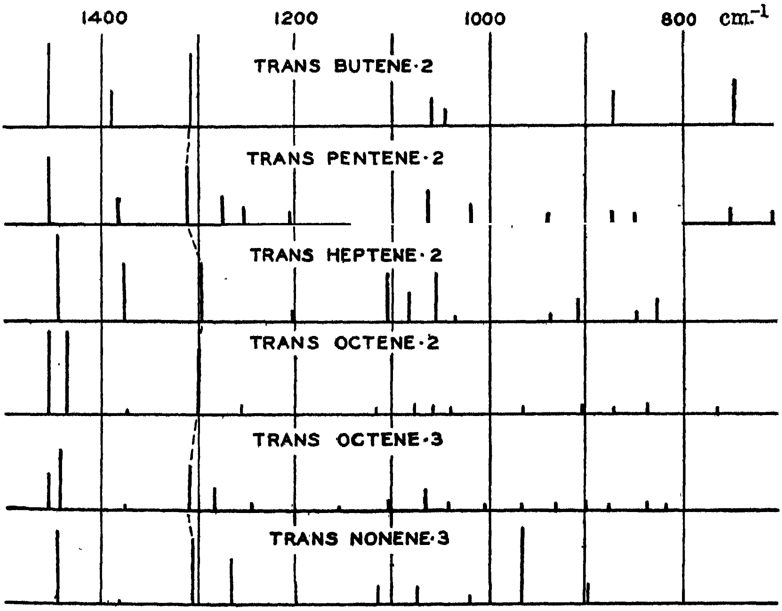


FIGURE 3. Raman spectra of *trans* disubstituted ethylenes.

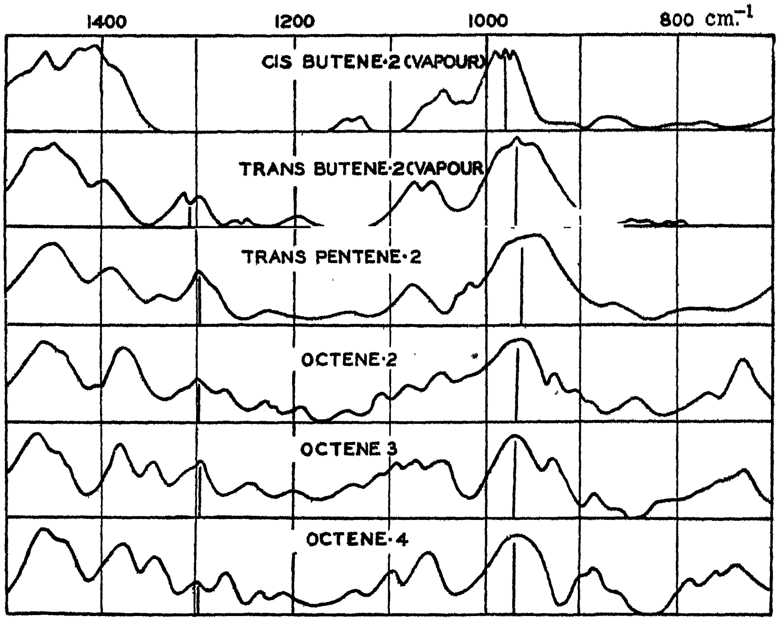
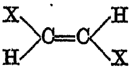
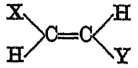
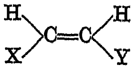
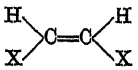


FIGURE 4. Infra-red spectra of *trans* and *cis* disubstituted ethylenes.

It should be observed that in the case of the more symmetrical molecule  $\begin{smallmatrix}X\\H\end{smallmatrix}\!>\!C{=}C\!<\!\begin{smallmatrix}H\\X\end{smallmatrix}$ there is a centre of symmetry (point group C_{2h}), so that infra-red active frequencies are forbidden in Raman spectra and vice-versa. Consequently with such molecules, there should be no coincidences between Raman and infra-red frequencies. However, the experimental data indicate that the frequency near 1300 cm.$^{-1}$ occurs with considerable intensity in both absorption and scattering even in the case of the simple molecule *trans*-butene-2. If the appearance of this frequency in both spectra had been a violation of the selection rules the intensities would have differed greatly in absorption and scattering. We accordingly assume that there are two distinct frequencies and assign them to two separate modes. Since they differ so slightly in numerical magnitude they must be either ν_1 and ν_2 or ν_3 and ν_4. Our previous arguments on the asymmetrically disubstituted molecule showed that the frequency of the bending mode in the plane (ν_1) was distinctly higher than that in which the hydrogens moved out of the plane (ν_3) and so we assign these two frequencies to ν_1 (Raman active) and ν_2 (infra-red active). The remaining strong infra-red band must then be ν_3. It should be noted that the weak Raman line in the region 965 to 980 cm.$^{-1}$ occurs only in the spectra of molecules of the type  $\begin{smallmatrix}X\\H\end{smallmatrix}\!>\!C{=}C\!<\!\begin{smallmatrix}H\\Y\end{smallmatrix}$ in which coincidences between infra-red and Raman frequencies are allowed. The Raman active ν_4 has not yet been identified.

Molecules of the type  $\begin{smallmatrix}H\\X\end{smallmatrix}\!>\!C{=}C\!<\!\begin{smallmatrix}H\\Y\end{smallmatrix}$ *and*  $\begin{smallmatrix}H\\X\end{smallmatrix}\!>\!C{=}C\!<\!\begin{smallmatrix}H\\X\end{smallmatrix}$

The assignment of the two observed C—H deformation frequencies in this case follows very naturally from the foregoing analysis of the corresponding *trans* form of molecule. The frequency near 1260 cm.$^{-1}$ (Raman, strong) is clearly to be associated with ν_1 of table 4 as this is the more symmetrical vibration and hence would be more intense in the Raman spectrum (figure 5). The 965 to 980 cm.$^{-1}$ band, being strong in absorption (figure 4) is most probably ν_3; ν_2 and ν_4 have not yet been identified. Although there is a fairly persistent Raman frequency near 1300 cm.$^{-1}$ in the spectra of these 'cis' disubstituted olefins (figure 5), the fact that it is absent in the spectra of *cis*-butene-2 and *cis*-pentene-2 has made us hesitate to assign it to ν_2. It should be added that it would be desirable to have the infra-red spectra of a few more *cis* compounds to confirm our assignment of ν_3.*

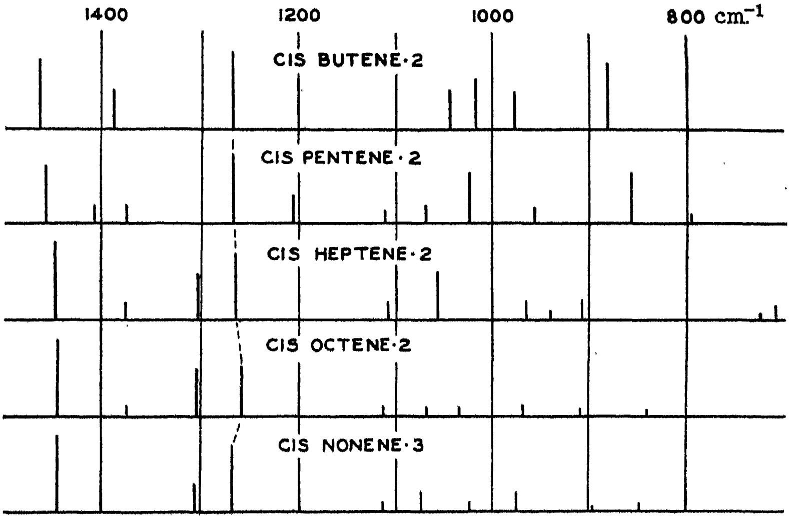
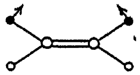
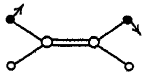
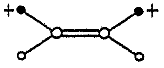
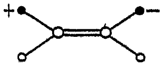


FIGURE 5. Raman spectra of *cis* disubstituted ethylenes.

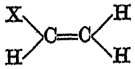
TABLE 4. THE ASSIGNMENT OF THE CHARACTERISTIC FREQUENCIES OF *CIS* DISUBSTITUTED ETHYLENES

designation	form of vibration	symmetry class	selection rules	assignment	intensity characteristics
ν_1		A_1	IR and R	1250–1270	R, strong
ν_2		B_1	IR and R	?	—
ν_3		B_2	IR and R	965– 980	IR, strong
ν_4		A_2	R	?	—

Having discussed the various disubstituted ethylenes we can now turn to the assignment of the deformation modes of the less symmetrical mono- and trisubstituted ethylenes.

---

* [*Note added in proof.*] Since the above was written, new infra-red data have become available on several *cis* compounds (American Petroleum Institute Reports, 1948) which indicate that ν_3 appears to lie near 690 cm.$^{-1}$ and not between 965 cm.$^{-1}$ and 980 cm.$^{-1}$.

Molecules of the type 

$$\mathrm{(X)(H)C{=}CH_2}$$

This skeleton belongs to the symmetry group C_s having nine modes involving motions in the plane of the model, and three out of the plane. Of the former nine modes, three must be essentially C—H deformation modes, while all of the out of plane motions must be hydrogenic. In contrast to the disubstituted ethylenes, even the approximate forms of vibration of the molecule cannot be drawn without a full normal co-ordinate treatment; accordingly we propose to label the planar modes ν_1, ν_2 and ν_3, and the non-planar modes as ν_1', ν_2' and ν_3', and to make assignments in the first instance to the ν or ν' type of vibration (table 5).

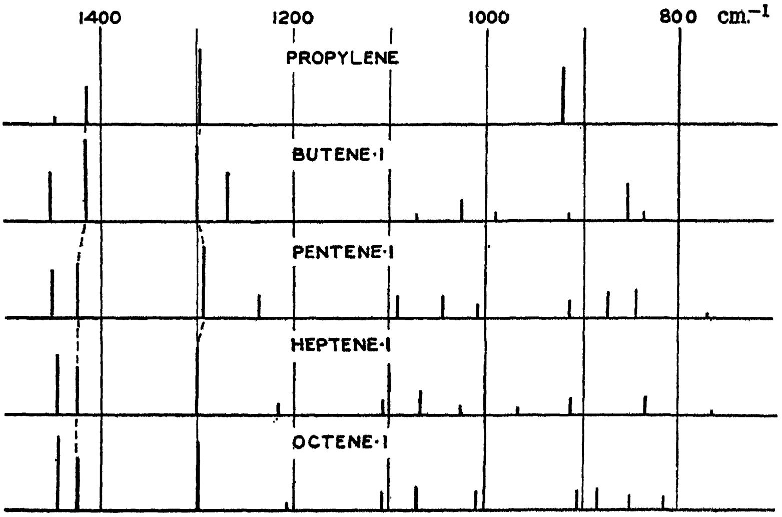


FIGURE 6. Raman spectra of singly substituted ethylenes.

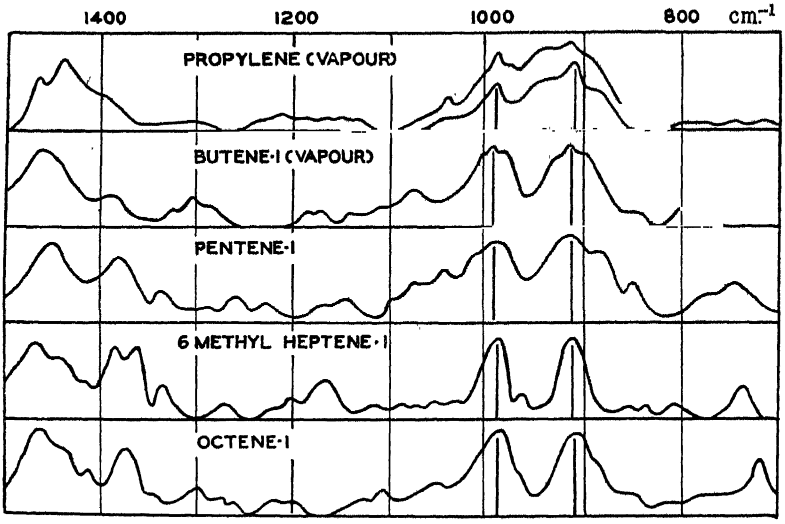


FIGURE 7. Infra-red spectra of singly substituted ethylenes.

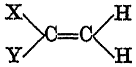
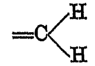
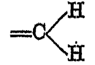
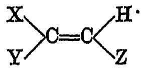
The experimental data (figures 6 and 7) provide four frequencies to be assigned among these different modes, viz. 1415, 1295, 990 and 910 cm.$^{-1}$. Following the above analysis of the more symmetrical molecules (since all four frequencies are close to those of the simpler disubstituted molecules) we note that two of them are strongly Raman active (1415 and 1295 cm.$^{-1}$) and obviously correspond respectively to the 1410 to 1420 cm.$^{-1}$ frequency of a  $\begin{smallmatrix}X\\Y\end{smallmatrix}\!\!>C=C<\!\!\begin{smallmatrix}H\\H\end{smallmatrix}$ skeleton and to the 1250 to 1300 cm.$^{-1}$ frequency found in both *cis* and *trans* disubstituted ethylenes. These are therefore assigned to planar vibrations, the first of which can be approximately described as a deformation vibration of the $=CH_2$ group, while the second is a similar vibration involving a hydrogen atom at each end of the double bond.

TABLE 5. THE ASSIGNMENT OF THE CHARACTERISTIC FREQUENCIES OF MONO-SUBSTITUTED ETHYLENES

designation	form of vibration	symmetry class	selection rules	assignment	strong in
ν_1	planar, mainly  $=C<\!\!\begin{smallmatrix}H\\H\end{smallmatrix}$	A'	IR and R	1410–1420	R
ν_2	planar, mainly $-CH{=}CH-$	A'	IR and R	1290–1300	R
ν_3	planar	A'	IR and R	?	—
ν_1'	non-planar, mainly  $=C<\!\!\begin{smallmatrix}H\\H\end{smallmatrix}$	A''	IR and R	905– 915	IR
ν_2'	non-planar, mainly $-CH{=}CH-$	A''	IR and R	985– 995	IR
ν_3'	non-planar	A''	IR and R	?	—

The intense infra-red bands at 905 to 915 and 985 to 995 cm.$^{-1}$ can similarly be compared with the non-planar vibrations of $=CH_2$ and $-CH{=}CH-$ groups which are only at slightly different frequencies, viz. 885 to 895 cm.$^{-1}$ and 964 to 980 cm.$^{-1}$ respectively, in the disubstituted ethylenes. The four observed frequencies can thus be assigned as in table 5 leaving ν_3 and ν_3', as yet, unidentified.

Molecules of the type  $\begin{smallmatrix}X\\Y\end{smallmatrix}\!\!>C=C<\!\!\begin{smallmatrix}H\\Z\end{smallmatrix}$

In this type of molecule there are only two deformation modes of the hydrogen atom, viz. (1) in the plane of the molecule, (2) out of the plane of the molecule. The experimental data (figures 8 and 9) and table 6 provide two characteristic frequencies. The strong infra-red band in the region 800 to 840 cm.$^{-1}$ can (by analogy with our previous assignments) be attributed to the non-planar mode. We believe that the variability in position of this frequency is at least partly due to changes in the precise nature of each of the alkyl substituents X, Y and Z which in turn can cause changes in the force constants controlling this vibration. In the mono- and disubstituted ethylenes there are fewer variables factors of this type.

The planar mode of the hydrogen is assigned to the Raman line near 1383 cm.$^{-1}$ having the correct magnitude for a planar vibration. It will be noted that this

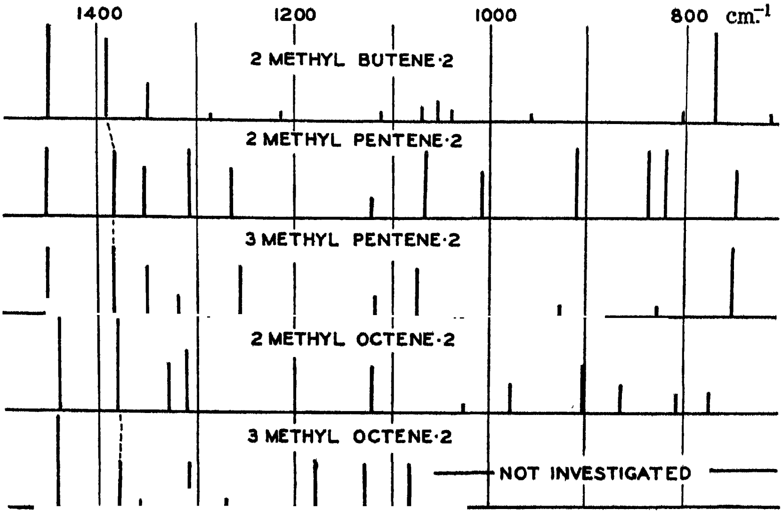


Figure 8. Raman spectra of trisubstituted ethylenes.

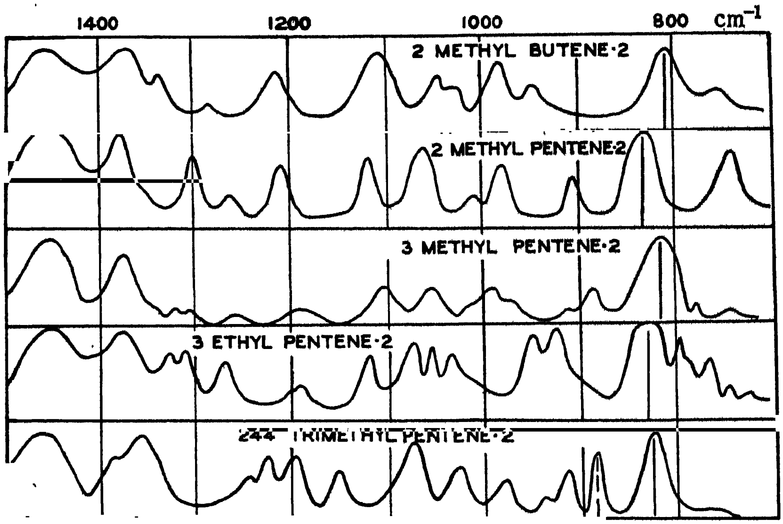
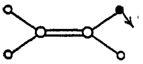
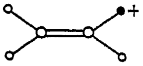


Figure 9. Infra-red spectra of trisubstituted ethylenes.

Table 6. The assignment of the characteristic frequencies of trisubstituted ethylenes

designation	form of vibration	symmetry class	selection rules	assignment	intense in
ν_1		A'	IR and R	1375–1390	R
ν_2		A''	IR and R	800– 840	IR

Raman frequency lies close to the well-known infra-red frequency near 370 cm.$^{-1}$, characteristic of methyl groups, and objection might be made to our assignment for this reason. We believe that such an objection is invalid since (1) the average frequency of this Raman line (1383 cm.$^{-1}$) is distinctly higher than the average for the infra-red frequency of methyl groups, and (2) the methyl group frequency is usually very weak in Raman scattering. This is well illustrated in the work of Fenske, Braun, Wiegand, Quiggle, McCormick & Rank (1947) on the Raman spectra of saturated hydrocarbons containing many methyl groups.

CALCULATIONS ON THE NON-PLANAR FREQUENCIES

Using a simple valency force field, the relations between the non-planar deformation frequencies of the hydrogen atoms, the masses of the various atoms, internuclear distances and angles have been derived. Two independent methods were used to give a check on the equations, the first being that of Howard & Wilson (1934), the second that of Eliashevich (1940). The general form of potential used was

$$2V = k_1\theta_1^2 + k_2\theta_2^2 + k_3\theta_3^2 + k_4\theta_4^2,$$

where θ_1 is the angle made by Mm_1 with the original plane of the ethylene skeleton when m_1 is displaced; θ_2, θ_3 and θ_4 are similarly defined (figure 10). The four θ's in this force-field are clearly not independent. However, this form of the potential function is easy to visualize and is very convenient for calculations on the more symmetrically substituted ethylenes. The required relation between the θ's is easily inserted in the course of the derivation of the relevant equations, or enters naturally in choosing the symmetry co-ordinates. When m_1, m_2, m_3 or m_4 is an alkyl group we assume that the latter may be replaced by a point mass; this is justified in that the deformation frequencies of the hydrogens will be shown to be very insensitive to these masses.

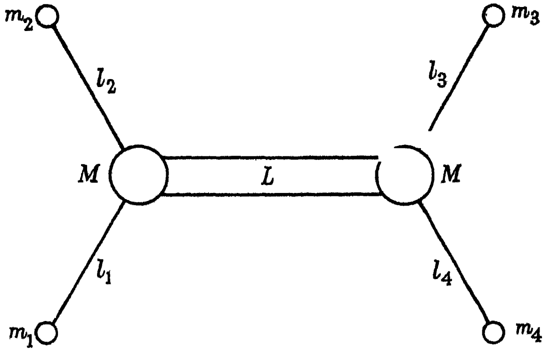


FIGURE 10

The relations obtained are as follows where $\lambda = 4\pi^2\nu^2$.

Asymmetrically disubstituted ethylenes (table 2)

$$\left.\begin{aligned} \lambda_3 + \lambda_3' &= kP + k'Q, \\ \lambda_3\lambda_3' &= kk'(PQ - R^2), \end{aligned}\right\} \tag{1}$$

where
$$P = 2\left(\frac{1}{M}\left(\frac{\cos\alpha}{L}+\frac{1}{l}\right)^2+\frac{\cos^2\alpha}{ML^2}+\frac{1}{2ml^2}\right)$$

and
$$Q = 2\left(\frac{1}{M}\left(\frac{\cos\alpha}{L}+\frac{1}{l'}\right)^2+\frac{\cos^2\alpha}{ML^2}+\frac{1}{2m'l'^2}\right)$$

and
$$R = \frac{2\cos\alpha}{ML}\left(\frac{2\cos\alpha}{L}+\frac{1}{l}+\frac{1}{l'}\right),$$

while
$$\lambda_4 = \frac{\dfrac{k}{m^2l^4}+\dfrac{k'}{m'^2l'^4}}{\dfrac{1}{ml^2}+\dfrac{1}{m'l'^2}}. \tag{2}$$

Here λ_3 and λ_4 correspond to ν_3 and ν_4 of table 2 while λ_3' is the counterpart of λ_3 in which non-hydrogenic atoms are deformed out of the plane. In this case

$$m_1 = m_2 = m', \quad m_3 = m_4 = m \quad \text{(hydrogen)},$$

$$l_1 = l_2 = l', \quad l_3 = l_4 = l \quad \text{(CH bond length)},$$

and
$$2V = k(\theta_3^2+\theta_4^2)+k'(\theta_1^2+\theta_2^2).$$

Trans *disubstituted ethylenes* (table 3)

$$\left.\begin{aligned}\lambda_3+\lambda_3' &= \frac{k}{l^2}\left(\frac{1}{m}+\frac{1}{M}\right)+\frac{k'}{l'^2}\left(\frac{1}{m'}+\frac{1}{M}\right),\\ \lambda_3\lambda_3' &= \frac{kk'}{l^2l'^2}\left(\frac{1}{mm'}+\frac{1}{mM}+\frac{1}{m'M}\right),\end{aligned}\right\} \tag{3}$$

$$\lambda_4 = \frac{k\left(\dfrac{S}{M}\left(\dfrac{2\cos\alpha}{L}+\dfrac{1}{l'}\right)+\dfrac{1}{ml^2}\right)^2+k'\left(\dfrac{S}{M}\left(\dfrac{2\cos\alpha}{L}+\dfrac{1}{l'}\right)+\dfrac{1}{m'l'^2}\right)^2}{\dfrac{S^2}{M}+\dfrac{1}{ml^2}+\dfrac{1}{m'l'^2}}, \tag{4}$$

where
$$S = \frac{4\cos\alpha}{L}+\frac{1}{l}+\frac{1}{l'}.$$

Here λ_3 and λ_4 correspond to ν_3 and ν_4 of table 3 while λ_3' is the counterpart of λ_3 in which the non-hydrogenic atoms are deformed out of the plane. In this case

$$m_1 = m_3 = m \text{ (hydrogen)}, \quad m_2 = m_4 = m',$$

$$l_1 = l_3 = l \text{ (CH bond length)}, \quad l_2 = l_4 = l',$$

and
$$2V = k(\theta_1^2+\theta_3^2)+k'(\theta_2^2+\theta_4^2).$$

Cis *disubstituted ethylenes* (table 4)

$$\left.\begin{aligned}\lambda_4+\lambda_4' &= kX+k'Y,\\ \lambda_4\lambda_4' &= kk'(XY-Z^2),\end{aligned}\right\} \tag{5}$$

where
$$X=\frac{1}{M}\left(\frac{2\cos\alpha}{L}+\frac{1}{l}\right)^2+\frac{1}{ml^2},$$
$$Y=\frac{1}{M}\left(\frac{2\cos\alpha}{L}+\frac{1}{l'}\right)^2+\frac{1}{m'l'^2},$$
$$Z=\frac{1}{M}\left(\frac{2\cos\alpha}{L}+\frac{1}{l}\right)\left(\frac{2\cos\alpha}{L}+\frac{1}{l'}\right),$$
$$\lambda_3=\frac{\dfrac{k}{l^2}\left(\dfrac{1}{ml}+\dfrac{1}{M}\left(\dfrac{1}{l}+\dfrac{1}{l'}\right)\right)^2+\dfrac{k'}{l'^2}\left(\dfrac{1}{m'l'}+\dfrac{1}{M}\left(\dfrac{1}{l}+\dfrac{1}{l'}\right)\right)^2}{\dfrac{1}{ml^2}+\dfrac{1}{m'l'^2}+\dfrac{1}{M}\left(\dfrac{1}{l}+\dfrac{1}{l'}\right)^2}. \tag{6}$$

In this case
$$m_1=m_4=m',\quad m_2=m_3=m\quad \text{(hydrogen)},$$
$$l_1=l_4=l',\quad l_2=l_3=l\quad \text{(CH bond length)},$$
and
$$2V=k(\theta_2^2+\theta_3^2)+k'(\theta_1^2+\theta_4^2).$$

General tests of the equations

Before applying the equations in detail to a series of experimentally determined frequencies, it is desirable to test whether they give rise to 'characteristic' frequencies under the appropriate conditions. If this proves to be true, it will afford additional evidence for our assignments and justify extending the treatment to other similar molecules.

In the asymmetrically disubstituted ethylenes, it has been shown that when the substituents are alkyl groups there is a characteristic frequency at 890 cm.$^{-1}$ which we attributed to the out-of-plane bending of the CH_2 group as a whole. The fact that this frequency is independent of whether the substituents are methyl, ethyl, propyl, etc. indicates that the frequency is independent of the precise masses, bond lengths and force constants associated with the substituent alkyl groups. The application of equations (1) to the asymmetrically disubstituted alkyl ethylenes yields the results given in table 7. α is assumed to be 120° throughout (as well known to be approximately correct for ethylene), $M=12$, $m=1$ (in atomic units) and $l=1{\cdot}07$ Å. The value of $k=0{\cdot}456\times10^{-11}$ ergs/radian was taken as a reasonable mean from the spectrum of ethylene (*vide infra*).

TABLE 7. RELATIVE VARIATIONS IN CERTAIN NON-PLANAR VIBRATIONS OF ASYMMETRICALLY DISUBSTITUTED ETHYLENES

m'	l'	$k\times10^{11}$	$k'\times10^{11}$	ν_3	ν_3'
1	1·07	0·456	0·456	1019	882
2	1·07	0·456	0·456	977	731
15	1·07	0·456	0·456	968	509
15	1·54	0·456	0·456	961	404
15	1·54	0·456	0·200	958	264

In the last two columns are given the calculated frequencies in wave numbers of the two normal modes of vibrations of this symmetry type and it can be seen that for m' greater than 2, changes of mass, bond length, or force constant for the substituent groups make very little difference to the ν_3 frequency. It is therefore legitimate to speak of this as a characteristic out-of-plane frequency of the CH_2 group in agreement with experiment. It is obvious that the value chosen for the out-of-plane bending constant of the C—H linkages is too high to obtain agreement with the observed band at 890 cm.$^{-1}$ and a more correct value will be given later.

A second test of the equations can be obtained by applying them to ethylene. For instance, putting $k = k'$, $m = m'$ and $l = l'$ in the asymmetrically disubstituted case, we find that equations (1) and (2) become

$$\left.\begin{aligned} \lambda_3 &= \frac{k}{l^2}\left(\frac{1}{m}+\frac{2}{M}\right), \\ \lambda_3' &= \frac{k}{l^2}\left(\frac{1}{m}+\frac{2}{M}\frac{(2l\cos\alpha+L)^2}{L^2}\right), \\ \text{and}\qquad \lambda_4 &= \frac{k}{l^2}\frac{1}{m}. \end{aligned}\right\} \tag{7}$$

The frequencies in ethylene corresponding to λ_3, λ_3' and λ_4 are respectively 949, 943 and 825 cm.$^{-1}$. These yield three independent values for k, viz. $0{\cdot}522\times10^{-11}$, $0{\cdot}392\times10^{-11}$ and $0{\cdot}460\times10^{-11}$ erg/radian. The fact that these are only in fair agreement with one another shows that the simple valency force field we have used, which neglects interaction terms, will have to be modified eventually when further data are available to allow these interaction terms to be estimated. Meanwhile, it is satisfactory to note that using the first two of the above separate values of k to calculate the corresponding two frequencies in tetradeuteroethylene, we obtain values of 717 and 780 cm.$^{-1}$ in excellent agreement with the experimental values of 720 and 780 cm.$^{-1}$ (Herzberg 1945). Thus, although the values we shall obtain for the various force constants by the use of these equations cannot be regarded as final, they should be very good first approximations and general trends in these, when derived from similar frequencies, can be regarded with some confidence, as reflecting general trends in the final values.

Application to asymmetrically disubstituted ethylenes

In addition to the hydrocarbons of this type whose spectra have already been discussed, various molecules with other substituents have been investigated by Raman and infra-red spectroscopy. Vinylidene fluoride, chloride and bromide have been examined in the infra-red by Thompson & Torkington (1945) and Torkington & Thompson (1945), the bromide in Raman scattering by de Hemptinne (1946), and the chloride in Raman scattering by Cabannes (1938).

Accepting our assignment of the strong band at 890 cm.$^{-1}$ in the olefines to the symmetrical CH_2 out-of-plane bending vibration, the assignments given in the literature for the same (B_1) vibrations of the halogen substituted ethylenes seem very reasonable. These are given in table 8 together with the force-constants deduced therefrom with the help of the appropriate equations. In carrying out these and subsequent calculations all planar angles are assumed to be 120° and the values in table 9 are assumed for the bond lengths. The values are taken from the work of Brockway (1936), Brockway, Beach & Pauling (1935), Gallaway & Barker (1942) and Huggill, Coop & Sutton (1938).

TABLE 8. TYPE B_1 HYDROGEN DEFORMATION FREQUENCIES OF ASYMMETRICALLY DISUBSTITUTED ETHYLENES

substituent, X	$\nu_3(B_1)$	$k \times 10^{11}$	k approx. $\times 10^{11}$
alkyl	890	0·394	0·403
bromine	883	0·394	0·398
chlorine	872	0·380	0·388
fluorine	801	0·320	0·328

TABLE 9. BOND LENGTHS IN SUBSTITUTED ETHYLENES

bond	length (Å)	bond	length (Å)
C—H	1·07	C—Cl	1·69
C=C	1·35	C—Br	1·83
C—CH_3	1·53	C—I	2·03
C—F	1·36		

In the above calculations generous upper and lower limits have been taken for the other out-of-plane frequency of type B_1 but these were found to make little difference to the value of k. An error of $\pm 0{\cdot}005$ in the above values of k will cover any uncertainty due to this cause. The 'k approx.' values were obtained by taking $\lambda_3 = kP$ as an approximate solution of the quadratic equation (1). This affords a very rapid means of approximating to the force constant direct from the vibration frequency and re-emphasizes the relative independence of the hydrogen deformation frequencies with regard to the structure at the other end of the double bond. It shows that any large change in frequency in passing from molecule to molecule must be mainly due to force constant variation.

In the case of vinylidene bromide, de Hemptinne (1946) gives the value of 709 cm.$^{-1}$ for the observed frequency of the deuterated molecule. The calculated value using the equations (1) and the value for the force constant given in table 8 is 703 cm.$^{-1}$, in good agreement with experiment. De Hemptinne (1946) also gives 668 cm.$^{-1}$ as the value of the twisting frequency of symmetry A_2 in the case of vinylidene bromide. This gives a value of $0{\cdot}301 \times 10^{-11}$ for k in our notation and is considerably less than $0{\cdot}394 \times 10^{-11}$ as given by the B_1 vibration. If the discrepancy is not due to a wrong assignment it means that there are considerable interaction constants present here also. In agreement with de Hemptinne's assignment, the former value of k gives a calculated value for the twisting vibration of the deuterated molecule of 473 cm.$^{-1}$ in good agreement with the experimental value of 478 cm.$^{-1}$. Thompson & Torkington (1946) give 540 cm.$^{-1}$ as the value of the torsional oscillation in vinylidene chloride. This seems too different from the vinylidene bromide value when the small change in the B_1 frequency in passing from the chloride to bromide is taken into account. More experimental data on the Raman spectra of molecules of this type are needed to settle these points.

Application to cis *and* trans *disubstituted ethylenes*

The experimental data on the spectra of halogen substituted molecules of this type are much less complete than for the corresponding asymmetrically disubstituted ethylenes, and the assignments are correspondingly less reliable. However, by starting from the well established frequency at 965 cm.$^{-1}$ for the alkyl substituted molecules and looking for intense infra-red bands (with marked Q branches) at lower frequencies it is possible to assign a fairly consistent set of frequencies to the symmetrical out-of-plane deformation vibrations of the C—H linkages. The assigned frequencies and the corresponding accurate and approximate force constants are given in tables 10 and 11.

TABLE 10. TYPE A_u HYDROGEN DEFORMATION FREQUENCIES OF *TRANS* DISUBSTITUTED ETHYLENES

substituent X	$\nu_3(A_u)$	$k \times 10^{11}$	k approx. $\times 10^{11}$
methyl	978	0·591	0·595
alkyl chain	965	0·575	0·579
iodine	895	0·496	0·498
chlorine	820	0·413	0·418

TABLE 11. TYPE B_2 HYDROGEN DEFORMATION FREQUENCIES OF *CIS* DISUBSTITUTED ETHYLENES

substituent X	$\nu_3(B_2)$	$k \times 10^{11}$	k approx. $\times 10^{11}$
methyl	986	0·633	0·640
alkyl chain	965	0·606	0·613
iodine	905	0·510	0·512
bromine	896	0·509	0·512
chlorine	860	0·474	0·477

The experimental data on the *cis* and *trans* dimethyl ethylenes (butenes) are taken from the infra-red work of Gershinowitz & Wilson (1938) and on the halogen substituted ethylenes from that of Wu (1937), Emschwiller & Lecomte (1937) and Herzberg (1945). The values chosen for the *cis* and *trans* dichloroethylenes are at variance with Wu's assignments. For example, the type A contour listed by Wu for the 860 cm.$^{-1}$ frequency of *cis*-dichloroethylene is assumed here to be of type C (both should have Q branches), and that at 820 cm.$^{-1}$ in the *trans*-compound is assumed to be the out-of-plane vibration rather than a C—Cl valence mode. These changes must be regarded as tentative only until more experimental work under high resolving power in the vapour phase has been done on these compounds. They are, however, in agreement with some recent assignments by Pitzer & Freeman (1946). Here again values are given for k as determined with the full equations and with the first term only, and the close correspondence of the two values in every case shows how separable are these particular modes. The k values may be taken as correct to $\pm 0{\cdot}005 \times 10^{-11}$ on the assumption of the simple valence force field and the stated molecular dimensions. The experimental data are insufficient to give reliable assignments to the other out-of-plane frequencies of symmetry types A_2 (*cis*) and B_g (*trans*), which will also be largely dependent on the force constants controlling the out-of-plane deformation of C—H linkages. The infra-red spectra of deuterated derivatives of these compounds do not appear to have been studied.

Application to the mono-substituted ethylenes

Experimental infra-red data on vinyl fluoride, chloride, bromide and iodide have recently been published by Torkington & Thompson (1945) and Thompson & Torkington (1945) and these, together with the data on mono-alkyl-substituted ethylenes given above form another interesting series of molecules in which to study frequency changes. It was pointed out earlier that the two strong infra-red frequencies of the mono-substituted ethylenes at approximately 910 and 990 cm.$^{-1}$ were respectively not far removed from the 890 and 965 cm.$^{-1}$ frequencies of asymmetrically disubstituted ethylenes. It was suggested that the two pairs of frequencies corresponded to closely related out-of-plane motions of the C—H linkages and it is therefore interesting to see if there is any parallel between the behaviour of these frequencies in going along the series methyl, alkyl, iodide, bromide, chloride, fluoride, in both cases.

It has already been shown that the out-of-plane deformation frequency of the $=CH_2$ group in asymmetrically disubstituted ethylenes is relatively independent of the masses, force constants, and bond lengths at the other end of the double bond. Also from tables 7 and 8 it is obvious that the change in frequency with X must be principally due to changes in the force constant k. If then the 910 cm.$^{-1}$ frequency of mono-alkyl substituted ethylenes is closely related to the 890 cm.$^{-1}$ frequency of the disubstituted compounds we should expect the frequency variation in the vinyl compounds to be considerably less than in the vinylidene compounds, since the effect of the substituent on k must be much less when there is one substituent than when there are two. This is borne out by table 12. In the second column is repeated the sequence of observed frequencies (from table 8) for the vinylidene compounds; the frequency differences between consecutive members are given in the third column. The corresponding figures for the vinyl compounds are given in columns four and five. It will be noticed that the differences in the fourth column ($\Delta\nu'$) are (within experimental error) the same or considerably less than the corresponding ones on the third column ($\Delta\nu$). Indeed if $\frac{1}{2}\Delta\nu$ is added to the observed value for vinyl fluoride and the corresponding operation carried out up the series we obtain 'calculated' frequencies in column 6 which lie remarkably close to the observed frequencies in column 5.

TABLE 12. COMPARISON OF RELATED DEFORMATION FREQUENCIES IN VINYL AND VINYLIDENE COMPOUNDS

	vinylidene compounds		vinyl compounds		
substituent	ν	$\Delta\nu$	$\Delta\nu'$	ν (obs.)	ν (calc.)
methyl	888	5	7	909	904
bromine	883	11	7	902	901
chlorine	872	71	35	895	896
fluorine	801			860*	860*

* Chosen equal to the experimental value.

A progressive change of this type along such a series provides additional strong evidence for our assignment of the 910 cm.$^{-1}$ band of the mono-substituted olefines. In the above compilation, the frequency assigned to this mode of vibration in vinyl fluoride (860 cm.$^{-1}$) is different from that of Torkington & Thompson (732 cm.$^{-1}$). However, the contour of the band is of the correct type and the manner in which the present assignment fits in with the above series is convincing. Another test of the hypothesis that each substituent makes its own contribution to the deformation force constant of the C—H linkage is afforded by the frequencies at 872, 835 and 801 cm.$^{-1}$ which are the out-of-plane vibrations of CH_2 groups in vinylidene chloride, 1-chloro-1-fluoro ethylene, and vinylidene fluoride (Torkington & Thompson 1945). The fact that this frequency in 1-chloro-1-fluoro compound is almost exactly the arithmetic mean of the other two is in agreement with the interpretation proposed.

Finally, de Hemptinne (1946) has given values for the fundamentals of vinyl bromide and its deuterated derivatives. He does not assign any frequencies specifically to out-of-plane modes but it is clear that his unassigned Raman frequency ω_8 (906 cm.$^{-1}$) in vinyl bromide is the out-of-plane mode of the CH_2 group. The infra-red value for this fundamental is given as 902 cm.$^{-1}$ by Thompson & Torkington (1945). De Hemptinne finds that the corresponding frequency in the completely deuterated compound is at 730 cm.$^{-1}$; calculations with our approximate equations give 724 cm.$^{-1}$. This close agreement gives further confirmation of the close relation between the 890 and 910 cm.$^{-1}$ frequencies in the alkyl substituted ethylenes.

DISCUSSION OF THE FORCE CONSTANTS

The question arises to what extent the force constants given in tables 8 to 11 provide a means of following the changes in molecular structure. Since these have been obtained on the assumption of a simple valence force field, they must be regarded as first approximations to a more complete set containing additional interaction constants. Nevertheless, it is reasonable to see what correlation obtains between them and the structural changes taking place as the substituents are varied. Except in the case of ethylene, the experimental data do not yet permit a more complete analysis. The validity of these approximate constants has been demonstrated in several instances in which they have been used to calculate the vibrational frequencies of the corresponding deuterium compounds. In cases where experimental data are available for such deuterated molecules the agreement between the calculated and experimental values has been within 2 %.

First it may be noted that among the force constants of the three types of dimethyl ethylenes as given in tables 8, 10 and 11, those corresponding to *cis* and *trans* molecules have fairly similar values (*ca.* $0{\cdot}6 \times 10^{-11}$ dynes cm./radian), whereas asymmetrically disubstituted molecules have a considerably lower value (*ca.* $0{\cdot}4 \times 10^{-11}$ dyne cm./radian). These differences may be in some measure due to differences in interaction constants in the different structures being reflected in the approximate values of the force-constants. It seems likely, however, that this does represent a real difference in the main deformation force constant. This same effect persists throughout the other sets of related molecules containing halogen atoms.

A second effect is common to the three series of force constants given in tables 8, 10 and 11. Here the substituents have been written down in the order of decreasing force constant, viz. methyl, alkyl group, iodine, bromine, chlorine, fluorine. This, however, is also the order of increasing electronegativity and we may conclude that there is strong evidence that the force constant controlling the out-of-plane bending of C—H linkages around substituted double bonds decreases with increasing electronegativity of the other substituents.

It is interesting to note that the force constants of the asymmetrically substituted ethylenes show relatively little change in going from methyl to bromine, and bromine to chlorine, but that there is a very considerable decrease in passing from chlorine to fluorine. This is in general agreement with the corresponding electronegativities according to Pauling (1941) where it is shown that the greatest change in electronegativity in this halogen series occurs in passing from chlorine to fluorine. It is also noticeable in proceeding along these series that for the *cis* and *trans* isomers, the change of force constant is much more rapid than for the asymmetrically disubstituted isomers. This is readily explained by the proximity of the substituent groups to the C—H linkages in question. In the asymmetrically disubstituted molecules, the substituents are at the end of the double bond remote from the C—H linkages, whereas in *cis* and *trans* isomers they are attached to the carbon atoms of the C—H linkages and the effects would be expected to be greater. These remarks apply equally to the variations of the 910 cm.$^{-1}$ frequency (the analogue of 890 cm.$^{-1}$ frequency in the asymmetrically disubstituted molecules) and 990 cm.$^{-1}$ frequency (the analogue of the 965 cm.$^{-1}$ frequency in the *cis* and *trans* isomers) in mono-substituted ethylenes.

It would seem to be profitable to develop the foregoing as a method of investigating changes in the electronic structure of the ethylenic double bond arising from substitution of the hydrogens around that bond. Changes in the force constants controlling the stretching vibrations of hydrogen atoms have already been correlated with changes in the electronic structure of these bonds (Sutherland & Dennison 1935; Linnett 1945) and it is therefore to be expected that variations in the deformation force constants of hydrogen atoms will receive a similar interpretation. Presumably there should be changes in the CH stretching force constant associated with the changes we have just discussed but unfortunately the spectroscopic data on the 3·3μ bands of substituted ethylenes have not been obtained under conditions of high enough resolving power to enable small changes in the CH stretching frequencies to be observed. Further experimental work along these lines is now required and we hope to make such observations in the near future. Furthermore, if the changes in the deformation force constants are due to changes in the electronic structure of the CH bonds, this should be reflected in changes in the absolute intensities of the associated absorption bands in the 8 to 13μ region of the spectrum as we proceed from iodine to fluorine. It is also hoped that further experimental work can be undertaken to check this prediction.

One of us (N.S.) is indebted to St Catharine's College, Cambridge, for a Senior Studentship and to the Dunlop Rubber Company for providing a maintenance grant, during the tenure of which this work was carried out. We also wish to thank Dr Rasmussen and Dr Brattain of the Shell Development Company for providing us in 1943 with advance information on their correlations of infra-red bands with the structure of substituted ethylenes (Rasmussen & Brattain 1947).

REFERENCES

American Petroleum Institute Research Project, **44**, 1947; Infra-red Spectrograms Nos. 19–23 contributed by the Shell Development Company, Emeryville, California, U.S.A.
Andant, A., Lambert, P. & Lecomte, J. 1935 *C.R. Acad. Sci., Paris*, **201**, 391.
Bourguel, M. 1931 *C.R. Acad. Sci., Paris*, **193**, 934.
Bourguel, M. 1932*a* *C.R. Acad. Sci., Paris*, **194**, 1736.
Bourguel, M. & Piaux, L. 1932*b* *Bull. Soc. Chim. Fr.* (IV), **51**, 1041.
Bourguel, M. 1933*a* *C.R. Acad. Sci., Paris*, **195**, 311.
Bourguel, M., Grèdy, B. & Piaux, L. 1933*b* *C.R. Acad. Sci., Paris*, **195**, 129.
Bourguel, M. 1935 *Bull. Soc. Chim. Fr.* (V), **2**, 1958.
Brockway, L. O. 1936 *Rev. Mod. Phys.* **8**, 231.
Brockway, L. O., Beach, J. Y. & Pauling, L. 1935 *J. Amer. Chem. Soc.* **57**, 2693.
Cabannes, J. 1938 *J. Chim. Phys.* **35**, 1.
Cleveland, F. F. 1943*a* *J. Chem. Phys.* **11**, 1.
Cleveland, F. F. 1943*b* *J. Chem. Phys.* **11**, 227.
Cleveland, F. F. 1943*c* *J. Chem. Phys.* **11**, 301.
Cleveland, F. F. 1944 *J. Chem. Phys.* **12**, 53.
Eliaschevich, M. A. 1940 *C.R. Acad. Sci. U.R.S.S.* **28**, 604.
Emschwiller, G. & Lecomte, J. 1937 *J. Phys. Rad.* **8**, 130.
Fellgett, P. B., Harris, G. P., Simpson, D. M., Sutherland, G. B. B. M., Thompson, H. W., Whiffen, D. H. & Willis, H. A. 1946 *Institute of Petroleum Reports.*
Fenske, M. R., Braun, W. G., Wiegand, R. V., Quiggle, D., McCormick, R. H. & Rank, D. H. 1947 *Industr. Engng Chem.* Anal. ed.), **19**, 700.
Fox, J. J. & Martin, A. E. 1940 *Proc. Roy. Soc.* A, **175**, 208.
Gallaway, W. S. & Barker, E. F. 1942 *J. Chem. Phys.* **10**, 88.
Gerhard, S. L. & Dennison, D. M. 1933 *Phys. Rev.* **43**, 197.
Gerschinowitz, H. & Wilson, E. B. Jr. 1938 *J. Chem. Phys.* **6**, 247.
Grèdy, B. 1935*a* *Bull. Soc. Chim. Fr.* (V), **2**, 1027.
Grèdy, B. 1935*b* *Bull. Soc. Chim. Fr.* (V), **2**, 1038.
Grèdy, B. 1935*c* *Bull. Soc. Chim. Fr.* (V), **2**, 1951.
Grèdy, B. 1937 *Bull. Soc. Chim. Fr.* (V), **4**, 415.
de Hemptinne, M. 1946 *Trans. Faraday Soc.* **42**, 5.
Herzberg, G. 1945 *Infra-red and Raman spectra of polyatomic molecules.* New York: Van Nostrand.
Howard, J. B. & Wilson, E. B. Jr. 1934 *J. Chem. Phys.* **2**, 630.
Huggill, J. A. C., Coop, I. E. & Sutton, L. E. 1938 *Trans. Faraday Soc.* **34**, 1518.
Lambert, P. & Lecomte, J. 1932 *Ann. Phys., Paris*, **18**, 329.
Lambert, P. & Lecomte, J. 1935 *C.R. Acad. Sci., Paris*, **201**, 391.
Lambert, P. & Lecomte, J. 1938*a* *C.R. Acad. Sci., Paris*, **206**, 1007.
Lambert, P. & Lecomte, J. 1938*b* *Ann. Phys., Paris*, **10**, 503.
Linnett, J. W. 1945 *Trans. Faraday Soc.* **41**, 223.
Kilpatrick, J. E. & Pitzer, K. S. 1947 *Bur. Stand. J. Res., Wash.*, **38**, 191.
Pauling, L. 1941 *The nature of the chemical bond.* Cornell University Press.
Pitzer, K. S. & Freeman, N. K. 1946 *J. Chem. Phys.* **14**, 586.
Rasmussen, R. S. & Brattain, R. R. 1947 *J. Chem. Phys.* **15**, 120, 131 and 135.
Simpson, D. M. & Sutherland, G. B. B. M. 1947 *J. Chem. Phys.* **15**, 153.
Sutherland, G. B. B. M. & Thompson, H. W. 1945 *Institute of Petroleum Reports.*
Sutherland, G. B. B. M. & Dennison, D. M. 1935 *Proc. Roy. Soc.* A, **148**, 250.
Thompson, H. W. & Torkington, P. 1945 *Proc. Roy. Soc.* A, **184**, 21.
Torkington, P. & Thompson, H. W. 1945 *Trans. Faraday Soc.* **41**, 236.
Wilson, E. B. Jr. & Wells, A. J. 1941 *J. Chem. Phys.* **9**, 319.
Wu, T. Y. 1934 *Phys. Rev.* **46**, 465.
Wu, T. Y. 1937 *J. Chem. Phys.* **5**, 392.

The intensity distribution in the nitrogen band systems emitted from the earth's upper atmosphere

By D. R. Bates, *University College, London*

(*Communicated by H. S. W. Massey, F.R.S.—Received* 30 *June* 1948
—*Read* 3 *March* 1949)

The intensity distributions in the negative and first and second positive band systems of nitrogen are determined for a number of excitation mechanisms. The results are compared with the intensity distributions observed from upper atmospheric sources with a view to understanding the processes occurring.

Ordinary aurorae are discussed in detail. It is shown that if the emission is attributed to electron excitation (as is commonly done) then it is necessary to assume that the vibrational temperature of the molecules is many thousands of degrees absolute—as similar views have been expressed by earlier writers on evidence that is now discredited it may be mentioned that in the present paper proper account *is* taken of the Frank-Condon principle. A more plausible interpretation of the observational results is that the heavy particles, that are often considered to form part of incident auroral streams, contribute to the excitation. High and low latitude aurorae show striking differences.

The emission of the negative band system from sunlit aurorae and from the twilight sky is studied. The direct action of the solar radiation on nitrogen ions is responsible. In the case of sunlit aurorae the continued absorption and emission of the resonance bands raises the vibrational temperature of the ions: this greatly influences the intensity distribution.

The consequences of the identification of the strong infra-red emission from the nocturnal sky with the (0, 0) band of the first positive system is discussed. It would appear that the identification necessitates a high degree of dissociation of the upper atmospheric nitrogen. Confirmatory observational work is urged.

In the appendix a preliminary investigation on the energy distribution of electrons in aurorae is attempted.

1. INTRODUCTION

In the study of the spectra of the nocturnal and twilight sky, and of aurorae, it is desirable that in addition to the accurate determination of the wave-length of each emission, attention should also be directed to the intensities, information on which is useful in establishing the validity of proposed identifications, and in determining the mechanisms that are responsible for the excitation.

There are two main aspects of the problem, the direct observational work, and the provision of basic comparison data. A fair amount of the former has already been done. It is with the latter that the present paper is concerned. The intensity distribution in various nitrogen band systems is investigated theoretically. The results obtained, supplemented by such laboratory measurements as are available, are used in a discussion of certain features of the radiations from the upper atmospheric sources.

2. THE METHOD OF CALCULATION OF RELATIVE PROBABILITIES OF THE TRANSITIONS BETWEEN THE VARIOUS VIBRATIONAL LEVELS OF DIFFERENT ELECTRONIC STATES OF NITROGEN

2·1. There is no difficulty in computing the relative probability of transitions between at least the lower vibrational levels of two electronic states. Both for transitions produced by electron impact and for those involving the absorption or emission of radiation the relative probability depends on the square of the modulus of the overlap integral,

$$p(n', v'; n'', v'') = \left| \int_0^\infty R(n', v' \mid r)\, R(n'', v'' \mid r)\, r^2\, dr \right|^2. \tag{1}$$

n' and n'' are here the quantum numbers describing the upper and lower electronic states respectively, v' and v'' are the vibrational quantum numbers; and $R(n', v' \mid r)$ and $R(n'', v'' \mid r)$ are the normalized vibrational wave functions, the variable r being the inter-nuclear distance. It is, of course, assumed that the value of the electronic transition integral does not depend on this last named quantity.

Neglecting certain minor terms the differential equation for the vibrational wave functions is simply

$$\left\{ \frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d}{dr} \right) + \frac{8\pi^2 \mu}{h^2} (\epsilon(n, v) - V(n \mid r)) \right\} R(n, v \mid r) = 0, \tag{2}$$

where μ is the reduced mass, h is Planck's constant, $\epsilon(n, v)$ the appropriate eigenvalue, and $V(n \mid r)$ the potential energy. Writing as usual

$$R(n, v \mid r) = r^{-1} S(n, v \mid r),$$

this becomes

$$\left\{ \frac{d^2}{dr^2} + \frac{8\pi^2 \mu}{h^2} (\epsilon(n, v) - V(n \mid r)) \right\} S(n, v \mid r) = 0. \tag{3}$$

A good representation of the potential energy is provided by the Morse function (1929),

$$\left.\begin{aligned} &D_e(n)\{1 - e^{-\delta(n)}\}^2, \\ \text{with} \quad &D_e(n) = hc\omega_e(n)/4x(n), \\ &\delta(n) = a(n)(r - r_e(n)), \\ \text{and} \quad &a(n) = \left\{ \frac{8\pi^2 \mu c}{h} \omega_e(n)\, x(n) \right\}^{\frac{1}{2}}, \end{aligned}\right\} \tag{4}$$

c being the velocity of light and $r_e(n)$, $\omega_e(n)$ and $x(n)$ having their standard interpretation in spectroscopic literature. Using (4) the solutions of (3) can, as is well known, be written down in analytical form in terms of the Laguerre polynomials. For $v = 0, 1, 2$ and 3 respectively the solutions are as follows:

$$S(n, 0 \mid r) = N(n, 0) \exp[-\{\delta(n)(1 - x_e(n)) + e^{-\delta(n)}\}/2x_e(n)], \tag{5}$$

$$S(n, 1 \mid r) = N(n, 1)\{S(n, 0 \mid r)/N(n, 0)\}\{1 - (1 - 2x_e(n))\, e^{\delta(n)}\}, \tag{6}$$

$$\begin{aligned} S(n, 2 \mid r) = N(n, 2)\{S(n, 0 \mid r)/N(n, 0)\}\{1 &- 2(1 - 3x_e(n))\, e^{\delta(n)} \\ &+ (1 - 3x_e(n))(1 - 4x_e(n))\, e^{2\delta(n)}\}, \end{aligned} \tag{7}$$

$$\begin{aligned} S(n, 3 \mid r) = N(n, 3)\{S(n, 0 \mid r)/N(n, 0)\}\{1 - 3(1 - 4x_e(n))\, e^{\delta(n)} \\ + 3(1 - 4x_e(n))(1 - 5x_e(n))\, e^{2\delta(n)} - (1 - 4x_e(n))(1 - 5x_e(n))(1 - 6x_e(n))\, e^{3\delta(n)}\}, \end{aligned} \tag{8}$$

the N's being normalizing factors. The molecular parameters $a(n)$, $x_e(n)$ and $r_e(n)$ have been found experimentally. Their values for the states $X\,{}^1\Sigma_g^+$, $A\,{}^3\Sigma_u^+$, $B\,{}^3\Pi_g$, $C\,{}^3\Pi_u$ of N_2 and $X\,{}^2\Sigma_g^+$, $B\,{}^2\Sigma_u^+$ of N_2^+ are shown in table 1, which is compiled from data given by Herzberg (1939) (for the ground states) and by Sponer (1935) (for the excited states). Sponer's figures were revised, using the more recent determinations of the basic physical constants, so as to make them consistent with those of Herzberg. Substituting from table 1, the first three or (in a few instances) four S functions* for the electronic states mentioned were evaluated and some 70 integrals of the type involved in (1) were computed numerically. $S(n,0)$, $S(n,1)$ and $S(n,2)$ are illustrated in figure 1 for the states of N_2, and N_2^+.

TABLE 1. MOLECULAR PARAMETERS

state	$a(n)$ (10^{-8} cm.$^{-1}$)	$x_e(n)$ (10^{-2})	$r_e(n)$ (10^{-8} cm.)
N_2 $X\,^1\Sigma_g^+$	2·449	0·612	1·095
$A\,^3\Sigma_u^+$	2·405	0·954	1·293
$B\,^3\Pi_g$	2·448	0·833	1·210
$C\,^3\Pi_u$	3·310	1·290	1·147
N_2^+ $X\,^2\Sigma_g^+$	2·588	0·731	1·117
$B\,^2\Sigma_u^+$	2·949	0·866	1·074

It is clear the S-functions are not sensitive to $a(n)$ or $x_e(n)$ so that the integrals depend mainly on $r_e(n)$ (which is probably well determined in all cases) and not on the details of the form of the potential energy curves. This suggests that the use of the Morse function is unlikely to introduce serious inaccuracies. It may be recalled too that Langstroth (1934, 1935) has evaluated a few of the p's, using as a model Hutchisson's harmonic oscillator (1930). His results generally do not differ greatly from those obtained by the more detailed treatment. For example, in the transition $C\,^3\Pi_u$–$X\,^1\Sigma_g^+$ of N_2 Langstroth gives $p(0,0)$, $p(1,0)$ and $p(2,0)$ to be 0·53, 0·34 and 0·08, whereas the present work gives them to be 0·52, 0·31 and 0·12. The errors caused by the assumption that the value of the electronic transition integral is independent of the inter-nuclear distance are difficult to estimate. Actually of course this integral is a function $i(r)$ of the inter-nuclear distance and the relative probabilities of transitions between the various vibrations depend not on the p's previously defined, but rather on quantities q given by,

$$q(n',v';n'',v'') = \left| \int_0^\infty i(r)\, R(n',v' \mid r)\, R(n'',v'' \mid r)\, r^2\, dr \right|^2. \tag{9}$$

As far as can be judged from the atomic or molecular orbital models, $i(r)$ does not in general vary at all rapidly for parallel-type transitions (i.e. $\Delta\Lambda = 0$)—usually less than as a first power of r. In order to obtain an indication of the importance of the effect a number of representative q's were computed taking $i(r)$ to be simply $k \times r$ and adjusting k the constant of proportionality so that $k \times \bar{r}_e$ is unity ($\bar{r}_e$ being the mean of the r_e's of the two states involved). Some of the values obtained, together with those of the corresponding p's are given in table 2. The differences are not unduly great. Those obtained if $i(r)$ is taken as k/r (with $k/\bar{r}_e$ unity) are of similar magnitude: but this is of less interest. The position regarding perpendicular-type transitions (i.e. $\Delta\Lambda = \pm 1$) is more uncertain.

* The higher vibrational levels were not treated as for these, the simple approximations used becomes less satisfactory.

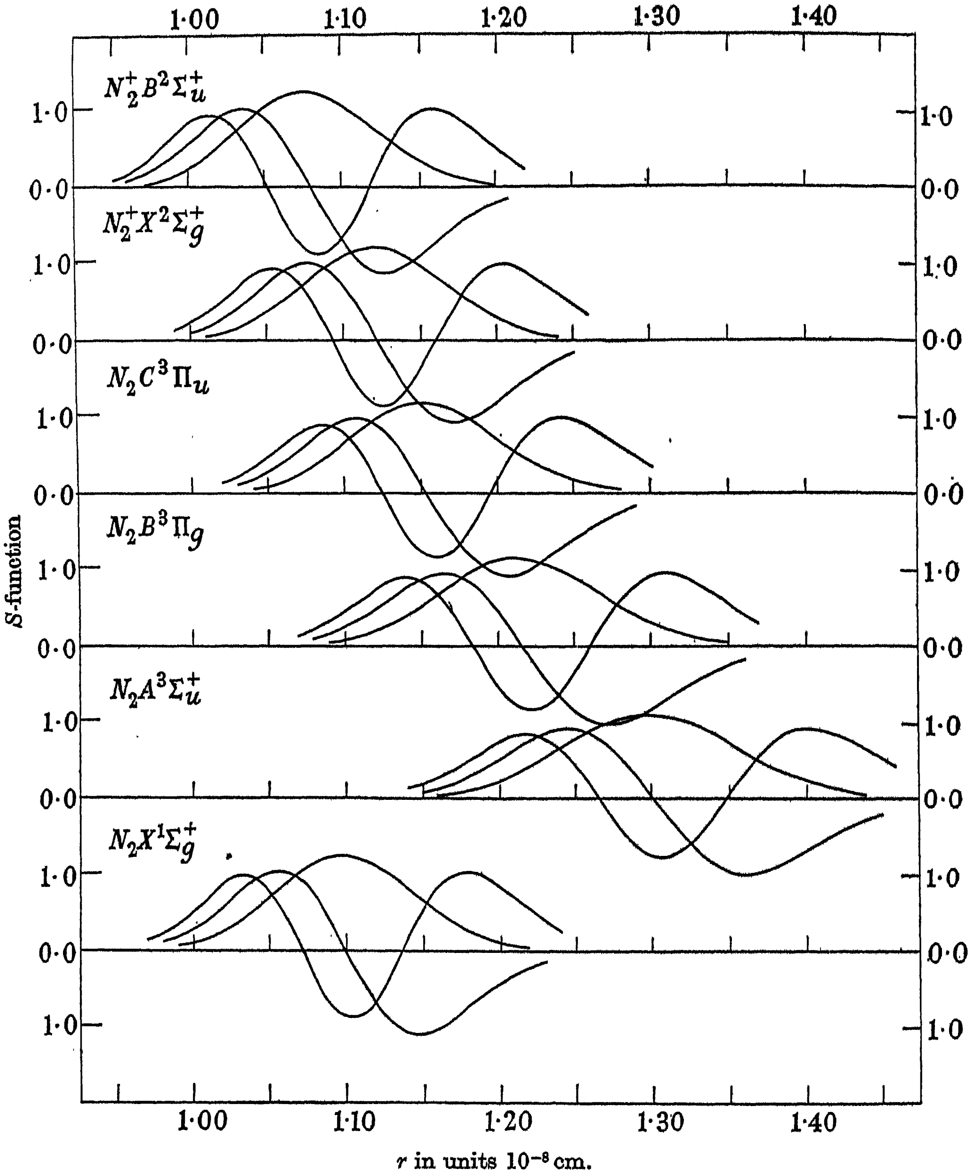


FIGURE 1. The first three S-functions of the various electronic states of N_2 and N_2^+. For normalization each must be multiplied by the factor $(\sqrt{8}\times 10^4)$ cm.$^{-\frac{1}{2}}$.

TABLE 2

transition	v'	v''	$p(n',v';n'',v'')$	$q(n',v';n'',v'')$
N_2^+, $B\,{}^2\Sigma_u^+$–N_2, $X\,{}^1\Sigma_g^+$	0	0	0·90	0·90
	2	0	0·0013	0·0023
N_2^+, $B\,{}^2\Sigma_u^+$–N_2^+, $X\,{}^2\Sigma_g^+$	0	1	0·26	0·24
	1	1	0·22	0·22
	1	2	0·29	0·34
	2	2	0·043	0·047

As will be seen later, the view that the theoretical results should be quite accurate is supported by comparison with experimental data. A particularly satisfactory feature is that there is no very marked worsening of the agreement for those bands that involve the higher vibrational levels studied (for which, of course, the theory is likely to be the least accurate).

2·2. Though the p's form a convenient basis for discussion it is usually not they but other physical quantities related to them that are of direct interest. The formulae involved are standard but are briefly stated here for reference.

Emission. The spontaneous transition probability associated with the band $(v'; v'')$* is, of course, proportional to $p(v', v'')\, \nu(v', v'')^3$, where $\nu(v', v'')$ is the frequency of the radiation emitted. If $g(v')$ is the rate of population of the levels v', the rate of photon emission in this band is therefore given by

$$a(v', v'') = g(v')\, p(v', v'')\, \nu(v', v'')^3 / \sum_{v''} p(v', v'')\, \nu(v', v'')^3 \text{ per sec.} \tag{10}$$

The intensity $I(v', v'')$ is naturally equal to $h\nu(v', v'')\, a(v', v'')$.

Excitation. If the number of molecules in the vibrational level v'' is $N(v'')$, the number of photons in the frequency range $\nu \to \nu + d\nu$ is $f_p(\nu)\, d\nu$, and the number of electrons in the energy range $\epsilon \to \epsilon + d\epsilon$ is $f_e(\epsilon)\, d\epsilon$, then the rates of excitation to the vibrational level v' due to the photons and electrons are respectively

$$\sum_{v''} \left\{ N(v'')\, p(v', v'') \int K_p(\nu) f(\nu)\, \nu\, d\nu \right\} \text{ per sec.} \tag{11}$$

and

$$\sum_{v''} \left\{ N(v'')\, p(v', v'') \int K_e(\epsilon) f(\epsilon)\, d\epsilon \right\} \text{ per sec.} \tag{12}$$

For simple excitation by photons, $K_p(\nu)$ is essentially zero unless ν is very close to $\nu(v', v'')$ when it can be taken to be a constant which is independent of v' and v''. If, as well as excitation, there is ionization (as $N_2, X\,{}^1\Sigma_g^+ \to N_2^+, B\,{}^2\Sigma_u^+$), $K_p(\nu)$ is zero only below the critical frequency: this is true also of $K_e(\epsilon)$ (either for simple excitation or for ionizing excitation): both can be regarded as approximately the same functions for all bands (*v. infra*, table 27, however).

Before proceeding to the discussion it may be mentioned that laboratory measurements usually only give the intensities of the various bands for *unknown* relative rates of population of the upper vibrational levels. Immediate comparison with theory is thus not possible. However, if the data are sufficiently complete, the p's can be obtained since, as can readily be seen,

$$p(v', v'') = \left\{ \frac{I(v', v'')}{\nu(v', v'')^4} \right\} \Big/ \left\{ \sum_{v''} \frac{I(v', v'')}{\nu(v', v'')^4} \right\}. \tag{13}$$

It is useful to note that when the experimental results are reduced in this way the following simple tests can be applied to check their accuracy.

(*a*) The sum of the p's in any v' progression should be unity,

$$\sum_{v'} p(v', v'') = 1. \tag{14}$$†

---

* For brevity the quantum numbers designating the electronic states will in future be omitted.

† The method of evaluating the p's of course ensures that the corresponding property of the v''-progressions is automatically satisfied.

(*b*) The sum of the square roots of the products of corresponding p's in any two v' or v'' progressions should, for some grouping of the signs, be zero,

$$\sum_{v'} \pm \{p(v', v''_A)\,p(v', v''_B)\}^{\frac{1}{2}} = \sum_{v''} \pm \{p(v'_C, v'')\,p(v'_D, v'')\}^{\frac{1}{2}} = 0 \quad \text{(when correct signs taken)} \qquad (15)$$

Both these tests can be deduced at once from the elementary properties of a complete set of normalized eigenfunctions. They depend, however, on the assumption that the electronic transition integral is approximately independent of the internuclear distance.

3. THE NEGATIVE SYSTEM

3·1. *The basic data*

The negative system arises from the $B\,{}^2\Sigma_u^+ \rightarrow X\,{}^2\Sigma_g^+$ transition of the molecular ion N_2^+. For convenience the wave-lengths of some of the principal members are quoted in table 3 which is based on the work of Herzberg (1928).

TABLE 3. WAVE-LENGTHS OF BANDS OF NEGATIVE SYSTEM (Å)

v' \ v''	0	1	2	3	4	5	6	7
0	3914	4278	4709	5228	5865			
1	3582	3884	4236	4652	5149	5754		
2	3308	3564	3858	4199	4600	5076	5653	
3	3078	3299	3549	3835	4167	4554	5012	5562
4		3076	3293	3538	3818	4141	4516	4958

Table 4 gives the p's for a number of the bands as calculated by the method described in § 2·1.

Herzberg (1928) has carried out an extensive series of intensity measurements in the laboratory. His results, converted to the form of p's are shown in table 5. For the $v' = 0$ and $v' = 1$ progressions the intensities of all but the feebler bands were measured so that the simple reduction procedure discussed in § 2·2 could be applied at once. Unfortunately a few of the bands of the other such progressions were obscured and in consequence their intensities could not be determined—in particular figures are not available for the bands (2, 0) and (3, 3). However, this difficulty can be at least partially overcome by making use of some of the properties of the p's mentioned in the previous section, and noting that the band system is confined mainly to low values of $|v' - v''|$. First, $p(2, 0)$ was chosen so that $\sum_{v'} p(v', 0) = 1$. The ratios of the remaining p's in the $v' = 2$ progression were obtained from the measurements as usual and their absolute magnitudes were fixed by the condition that $\sum_{v''} p(2, v'') = 1$ (with $p(2, 0)$ as just chosen). For the $v'' = 3$ progression the measurements again enabled the ratio of all the p's except one to be determined and their absolute magnitudes were found by adjusting $p(3, 1)$ so that $\sum_{v''} p(v'', 1) = 1$. Finally the relation $\sum_{v''} p(3, v'') = 1$ gave the missing $p(3, 3)$. The indirectness of this method of analysis probably does not introduce very much uncertainty. It will be noted that the p's given satisfy test (b) of § 2·2 in as far as it can be applied. Thus

$$+(p(0,0)\,p(0,1))^{\frac{1}{2}}-(p(1,0)\,p(1,1))^{\frac{1}{2}}-(p(2,0)\,p(2,1))^{\frac{1}{2}}-(p(3,0)\,p(3,1))^{\frac{1}{2}} = 0{\cdot}00$$

as is required, and

$$+(p(0,0)\,p(0,2))^{\frac{1}{2}}-(p(1,0)\,p(1,2))^{\frac{1}{2}}+(p(2,0)\,p(2,2))^{\frac{1}{2}}+(p(3,0)\,p(3,2))^{\frac{1}{2}} = -0{\cdot}02,$$

which is only slightly in error. Test (a) cannot, of course, be applied as it was used in the building up of the table.

Ornstein & van Wijk (1928) have measured the relative intensities of a small number of the bands. Their results are sufficient to enable $p(0,0)$ and $p(0,1)$ to be estimated as about 0·69 and 0·23 respectively (in close agreement with Herzberg). Crude lists of intensities have been published by Merton & Pilley (1925) and by Smyth & Arnott (1930) but these, though useful in confirming the general trend, are not precise enough to be used in quantitative discussion.

TABLE 4. p's FOR N_2^+, $B\,^2\Sigma_u^+ - X\,^2\Sigma_g^+$ TRANSITION (THEORY)

$v'' \backslash v'$	0	1	2
0	$0{\cdot}65$	$0{\cdot}26$	$0{\cdot}07_1$
1	$0{\cdot}30$	$0{\cdot}22$	$0{\cdot}29$
2	$0{\cdot}04_7$	$0{\cdot}41$	$0{\cdot}04_3$
3	$0{\cdot}00_3$	$0{\cdot}11$	$0{\cdot}41$

TABLE 5. p's FOR N_2^+, $B\,^2\Sigma_u^+ - X\,^2\Sigma_g^+$ TRANSITION (LABORATORY MEASUREMENT)

$v'' \backslash v'$	0	1	2	3	4	5	6	7
0	$0{\cdot}69$	$0{\cdot}26$	$0{\cdot}04_8$	$0{\cdot}00_3$	$0{\cdot}00_0$			
1	$0{\cdot}23$	$0{\cdot}27$	$0{\cdot}38$	$0{\cdot}11$	$0{\cdot}01_3$	$0{\cdot}00_3$		
2	$(0{\cdot}07_5)$	$0{\cdot}39$	$0{\cdot}07_2$	$0{\cdot}25$	$0{\cdot}18$	$0{\cdot}02_7$	$0{\cdot}00_6$	
3	$0{\cdot}00_1$	$0{\cdot}08_0$	$0{\cdot}36$	$(0{\cdot}03_5)$	$0{\cdot}20$	$0{\cdot}27$	$0{\cdot}04_3$	$0{\cdot}01_1$

Comparison of tables 4 and 5 shows that there is fair accord between the p's obtained by calculation and those obtained from the laboratory measurements. It seems justifiable therefore to regard the values with some confidence. The most convenient final way of presenting the results for the future applications is to give them in the form of the band intensities that would ensue if there were equal rates of population of the vibrational levels of the excited state (cf. § 2·2). This is done in table 6. The figures are an average of those derived from the two sets of p's. The units are arbitrary.*

In addition to their importance so far as the actual emission is concerned, tables 4 and 5 are of value in studying excitation from the ground state of the ionized molecule. So that excitation from the ground state of the neutral molecule can be discussed, calculations were performed on the p's for the transition $\mathrm{N_2}, X\,^1\Sigma_g^+ \rightarrow \mathrm{N_2^+}, B\,^2\Sigma_u^+$ and (less important but desirable for completeness) $\mathrm{N_2}, X\,^1\Sigma_g^+ \rightarrow \mathrm{N_2^+}, X\,^2\Sigma_g^+$. Tables 7 and 8 contain the results derived.

---

* Actually they were chosen so that for the $v' = 0$ progression the intensities obtained from the theoretical p's are simply given by $100p(0,v'')\left\{\dfrac{\nu(0,v'')}{\nu(0,0)}\right\}^4$. These units (or similar ones for other band systems) will be used throughout the paper.

Table 6. Predicted intensity distribution of negative bands assuming equal rates of population of the vibrational levels of the excited state

v' \ v''	0	1	2	3	4	5	6	7
0	66	18	2·8	0·1	0·0			
1	35	24	23	5·2	0·4	0·1		
2	10	50	5·2	16	8·1	0·9	0·1	
3	0·4	16	48	3·3	14	13	1·4	0·2

Table 7. p's for $\mathrm{N_2^+}, B\,^2\Sigma_u^+ – \mathrm{N_2}, X\,^1\Sigma_g^+$ (theory)

v' \ v''	0	1	2
0	0·90	$0{\cdot}08_8$	$0{\cdot}00_9$
1	$0{\cdot}09_5$	0·74	0·14
2	$0{\cdot}00_1$	0·17	0·63

Table 8. p's for $\mathrm{N_2^+}, X\,^2\Sigma_g^+ – \mathrm{N_2}, X\,^1\Sigma_g^+$ (theory)

v' \ v''	0	1	2
0	0·89	0·11	$0{\cdot}00_4$
1	0·10	0·68	0·20
2	$0{\cdot}00_8$	0·18	0·50

Enough has now been done to enable the intensity distribution arising from a variety of mechanisms and conditions to be determined rapidly.

Two obvious cases at once present themselves, excitation by the collision of electrons with neutral or ionized molecules in their ground electronic states and in their zeroth vibrational levels. Using equations (10) and (12) (with $f(\epsilon)$ taken as constant) tables 6 and 7, and the mean of tables 4 and 5, the intensities given in tables 9 and 10 were obtained. In examining these it should be noted that the same *total* excitation rate was assumed for both.

A useful partial check on the correctness of table 9 (and in consequence of table 7) is provided by the work of Langstroth (1934) who carried out an experiment in which no excitation process, other than the one there considered, could be appreciable. He found the ratio of the intensities of the (0, 1) and (1, 2) bands to be 7·0. As can be seen from the figures in the table 7·3 is the corresponding predicted ratio. The agreement is thus excellent. It is perhaps worth mentioning that Langstroth was able to verify that the ratio does not depend on the electron energy even when this is only 21 eV (the lowest employed).

No experimental data exist with which to compare the results given in table 10. However, their essential correctness cannot be in doubt as they depend only on the value of the p's associated with the N_2^+, $B\,^2\Sigma_u^+-X\,^2\Sigma_g^+$ transition which as noted earlier seem to be well established.

TABLE 9. PREDICTED INTENSITY DISTRIBUTION OF NEGATIVE BANDS AS EXCITED BY THE COLLISION OF ELECTRONS WITH NEUTRAL NITROGEN MOLECULES IN THEIR GROUND ELECTRONIC STATE AND ZEROTH VIBRATIONAL LEVEL

v' \ v''	0	1	2	3
0	59	16	2·5	0·1
1	3·3	2·2	2·2	0·5
2	0·0	0·1	0·0	0·0

TABLE 10. PREDICTED INTENSITY DISTRIBUTION OF NEGATIVE BANDS AS EXCITED BY THE COLLISION OF ELECTRONS WITH IONIZED NITROGEN MOLECULES IN THEIR GROUND ELECTRONIC STATE AND ZEROTH VIBRATIONAL LEVEL

v' \ v''	0	1	2	3	4
0	44	12	1·9	0·0	0·0
1	9·4	6·3	6·1	1·4	0·1
2	0·6	3·1	0·3	1·0	0·5
3	0·0	0·0	0·1	0·0	0·0

TABLE 11. RELATIVE RATES OF POPULATION OF THE VARIOUS VIBRATIONAL LEVELS OF N_2^+, $B\,^2\Sigma_u^+$ BY ELECTRON IMPACT WITH (*a*) N_2, $X\,^1\Sigma_g^+$ AND (*b*) N_2^+, $X\,^2\Sigma_g^+$

effective vibrational temperature (° K) ...		500	1000	2000	4000	6000	8000
case (*a*)	$v' = 0$	1·00	1·00	1·00	1·00	1·00	1·00
	1	0·11	0·13	0·26	0·47	0·60	0·68
	2	0·00	0·01	0·07	0·22	0·36	0·47
case (*b*)	$v' = 0$	1·00	1·00	1·00	1·00	1·00	1·00
	1	0·40	0·41	0·46	0·57	0·66	0·71
	2	0·09	0·12	0·21	0·36	0·47	0·55

Under upper atmospheric conditions the particles are not necessarily confined to their zeroth vibrational level and it is desirable to investigate the effect of this. The calculation of the relative rates of population of the various vibrational levels of the excited state was therefore carried out as before but with a Boltzmann distribution among the vibrational levels of the ground state. Table 11 gives the results for a wide temperature range. From these and table 6 the relative intensities of the bands can be deduced immediately. The expression for the corresponding rates for excitation by sunlight is given in equation (11). Unfortunately, nothing is known about the function $f_p(\nu)$ for the ionization process

$$N_2(X\,^1\Sigma_g^+) + h\nu \rightarrow N_2^+(B\,^2\Sigma_u^+) + e, \qquad (16)$$

as the photons involved lie outside the observable spectral region. If, however, the energy flux can be assumed to be constant over the relatively small frequency range of importance the same figures as those in table 11 (case (*a*)) for excitation by electron bombardment can be taken. It is possible to treat with greater precision the more interesting direct resonance process

$$N_2^+(X\,^2\Sigma_g^+) + h\nu \rightarrow N_2^+(B\,^2\Sigma_u^+). \tag{17}$$

The relevant $f(\nu)$ was found by combining the solar radiation measurements of Minnaert, Mulders & Houtgast (1940) and of Pettit (1940) and integrating over the individual bands (taking account of their rotational structure). In table 12 the relative population rates obtained are given.

TABLE 12. RELATIVE RATES OF POPULATION OF THE VARIOUS VIBRATIONAL LEVELS OF N_2^+, $B\,^2\Sigma_u^+$ BY THE ACTION OF SOLAR RADIATION ON N_2^+, $X\,^2\Sigma_g^+$

effective vibrational temperature (° K) ...	500	1000	2000	4000	6000	8000
$v' = 0$	1·00	1·00	1·00	1·00	1·00	1·00
1	0·18	0·19	0·24	0·38	0·48	0·55
2	0·05	0·06	0·12	0·25	0·35	0·43

Before proceeding to the study of the upper atmospheric emission one further mechanism that may influence the relative intensities must be considered. So far in discussing the excitation of the negative bands system from N_2^+, $X\,^2\Sigma_g^+$, by electron impact or by photo-absorption, it has been assumed that the distribution among the vibrational levels is fixed in some specified way. Actually, however, unless collisions are very effective in maintaining it the distribution is rapidly altered since changes in the vibrational quantum number may occur in the primary transition

$$N_2^+,\ X\,^2\Sigma_g^+ \rightarrow N_2^+,\ B\,^2\Sigma_u^+$$

and in the subsequent transition

$$N_2^+,\ B\,^2\Sigma_u^+ \rightarrow N_2^+,\ X\,^2\Sigma_g^+.^*$$

In the equilibrium that is ultimately established the gain and loss to any vibrational level must clearly be equal. The resultant distribution in the extreme case where activating or deactivating collisions are ignored is therefore governed by the following set of simultaneous equations

$$\sum_{v'_k}\sum_{v''_j} n(v''_j)\,\alpha(v''_j, v'_\kappa)\,\beta(v'_k, v''_i) = n(v''_i)\sum_{v'_j}\alpha(v''_i, v'_j) \quad (\text{all } v''_i). \tag{18}$$

Here the n's denote the equilibrium relative number of ions in each vibrational level, the α's the excitation rates/unit concentration, and the β's the photon emission/unit excitation rate: as always v' and v'' respectively indicate the vibrational quantum numbers in the upper and lower electronic states. In the notation of § 2·2

---

* The corresponding effect for processes such as (16) is unimportant.

$$\left.\begin{aligned}\alpha(v'',v') &= p(v',v'')\int K_e(\epsilon)f(\epsilon)\,d\epsilon,\\ \text{or}\qquad &= p(v',v'')\int K_p(\nu)\,f(\nu)\,\nu\,d\nu\end{aligned}\right\} \tag{19}$$

and

$$\beta(v',v'') = p(v',v'')\,\nu(v',v'')^3/\sum_{v''} p(v',v'')\,\nu(v',v'')^3. \tag{20}$$

Using these relations and the data already tabulated, the coefficients appearing in (18) were evaluated and the equations solved to give the n's. The rates of population of the vibrational levels of N_2^+, $B\,^2\Sigma_u^+$ were then determined as usual, the results obtained being given in table 13. For simplicity those levels with vibrational quantum number equal to or greater than 3 were throughout grouped together.

TABLE 13. RELATIVE RATES OF POPULATION OF THE VARIOUS VIBRATIONAL LEVELS OF N_2^+, $B\,^2\Sigma_u^+$ THROUGH THE EXCITATION OF N_2^+, $X\,^2\Sigma_g^+$ (IN THE SPECIAL EQUILIBRIUM DISTRIBUTION) (*a*) BY ELECTRON IMPACT, (*b*) BY THE ACTION OF SOLAR RADIATION

v'	case (*a*)	case (*b*)
0	1·00	1·00
1	0·80	0·40
2	0·69	0·27
$\geqslant 3$	(0·63)	(0·15)

3·2. *Discussion*

The observational data on the relative intensities of the members of the negative system as emitted from an upper atmospheric source is still far from complete so that full use cannot as yet be made of the results compiled in the preceding section. A preliminary discussion can, however, be attempted.

3·2·1. Perhaps the most accurate measurements that have been made are those on low latitude aurorae.* Working in Haute Provence, Barbier (1947) was able to obtain a spectrum of the great aurora that occurred, during the nights of 17 and 18 April 1947. He estimated that the altitude of the aurora was probably of order 250 to 300 km. and judged that it was not illuminated by the sun.† He found that the negative system showed strongly with an intensity distribution very different from that generally accepted as characteristic of ordinary aurorae such as are observed in Norway (*v. infra*). The relative intensities (corrected for atmospheric extinction) he obtained are quoted in table 14. They are given to two figures for convenience, but it should be mentioned that such accuracy is not claimed by Barbier who actually tabulated their logarithms, and these he only gave to within

---

* These aurorae show certain very interesting features and further detailed study of them seems likely to yield valuable information on the upper atmosphere (cf. review of Barbier (1948)).

† The details regarding the aurora not given in the original publication were kindly supplied to me by Professor Barbier in a private communication.

0·05 (which corresponds to giving the relative intensity itself to within about one part in ten). The first aspect of the distribution to consider is the variation along a v'' progression. Though this does not of course give any information about the excitation process that is involved it is of importance in that it serves as a guide, to the accuracy of the measurements, and to the extent to which they can be relied upon to represent the true relative intensities *as emitted*. The most instructive way of proceeding is to derive the p's. In spite of the fact that data are lacking for a number of the bands this can be done fairly satisfactorily by adjusting the absolute scale so that the sums of certain suitably chosen groups of the p's equal the corresponding sums based on the theoretical and laboratory data. Table 15 gives the results obtained. Taking account of the difficulties associated with auroral observations the agreement with tables 4 and 5 is remarkably close. This suggests that Barbier's measurements can be treated as trustworthy.

TABLE 14. INTENSITY DISTRIBUTION OF NEGATIVE SYSTEM AS OBSERVED IN LOW LATITUDE AURORAE

(0 indicates band not recorded; — indicates band outside spectral range apparently studied)

v' \ v''	0	1	2	3	4
0	32	7·9	—	—	—
1	8·9	10	5·6	—	—
2	—	8·9	0	4·0	—
3	—	—	7·9	0	0

TABLE 15. p'S FOR N_2^+, $B\,^2\Sigma_u^+ – X\,^2\Sigma_g^+$ TRANSITION (LOW LATITUDE AURORAE)

(In some cases it is only possible to say whether the p value is large (l) or small (s))

v' \ v''	0	1	2	3	4
0	0·68	0·25	—	—	—
1	0·22	0·34	0·28	—	—
2	—	0·34	s	0·29	—
3	—	—	l	s	(s)

Turning next to the intensity distribution along a v' progression to see what evidence this provides regarding the excitation it is at once apparent by comparison of table 14 with tables 9 and 10 that the simplest processes that can be postulated, electron collision with either N_2, $X\,^1\Sigma_g^+$ or N_2^+, $X\,^2\Sigma_g^+$ in their zeroth vibrational level, (which is essentially the equilibrium distribution at low temperature) are unacceptable as they lead to an insufficiently developed band system. Other more complex excitation conditions have therefore to be examined. This can most readily be done, not through the consideration of the relative intensities themselves, but rather through the consideration of $g(v')$, the relative rates of population of the various vibrational levels of the upper state (which can be derived from them by using table 6). For this purpose table 16 was compiled: the figures given in it were obtained by averaging over all the bands in each progression. The regularity of the variation of $g(v')$ is such as to suggest that the general trend exhibited is of real significance. However, the slowness of the fall-off is so remarkable that some confirmation of it is desirable. The examination of suitable spectra for members of the $v' = 4$ progression would be valuable in this connexion. It is perhaps useful to mention here that the strongest members are probably (4, 2) $\lambda 3293$ and (4, 3) $\lambda 3538$ and that crude estimates* indicate that the intensities of both these (in the units of table 14) should be about 3 or 4.

TABLE 16. RELATIVE RATES OF POPULATION OF THE VARIOUS VIBRATIONAL LEVELS OF N_2^+, $B\,^2\Sigma_u^+$ (LOW LATITUDE AURORAE)

v'	0	1	2	3
$g(v')$	1·00	0·65	0·46	0·34

Comparison of tables 16 and 11 shows once again that the auroral emission cannot arise from the bombardment, by electrons, of either N_2, $X\,^1\Sigma_g^+$ or N_2^+, $X\,^2\Sigma_g^+$ if their vibrational temperature $\Theta(v)$ is low. Further, though there is better agreement between the relative population rates in the two tables at high $\Theta(v)$ it is not fully satisfactory except at about 8000° K, for excitation from the neutral molecule, or at about 6000° K, for excitation from the ionized molecule. Both these values are very high and neither could be accepted as applying to the translational temperature $\Theta(t)$.† It is, of course, conceivable that $\Theta(v)$ and $\Theta(t)$ are not equal but that instead the former is very much larger than the latter.

As far as the neutral molecules are concerned there is, however, no obvious action that seems capable of supplying vibrational energy sufficiently rapidly to overcome the usual processes tending to produce thermal equilibrium: for example, neither the yield from the association of free atoms to form vibrationally excited molecules nor the action of radiation (absorption followed by emission with overall increase in vibrational energy) appear likely to be adequate. The necessity for invoking a high $\Theta(v)$ for the neutral molecules might be avoided if an appreciable fraction of the excitations arose from the action

$$N_2(A\,^3\Sigma_u^+) + e \rightarrow N_2^+(B\,^2\Sigma_u^+) + 2e. \tag{21}$$

Though detailed calculations have not been performed it is clear from the magnitude of the change in the equilibrium internuclear distance (0·219 Å) that this contributes mainly to the population rates to the upper vibrational levels (see figure 1). There is little doubt, however, that the concentration in the metastable state is so low that process (21) is in fact unimportant.

---

* Based on $p(4, 2) = 0{\cdot}16$ and $p(4, 3) = 0{\cdot}26$ (values derived from Herzberg's measurements as in § 3·1 but subject to some uncertainty because of lack of completeness of the data): and on $g(4) = 0{\cdot}25$ (a dubious extrapolation).

† Cf. the survey by Spitzer (1947). Confirmatory work is desirable to establish that inaccuracies in the intensity data do not cause the temperature required to be seriously overestimated. The extent of the overestimate possible is difficult to judge but it seems unlikely that it can be such as to allow a value of even 4000° K. In the 250 to 300 km. region (where the aurora under discussion is apparently located) the temperature is usually considered to be 2000° K or less.

The position regarding the ionized molecules is interesting as these tend automatically to become vibrationally excited as a consequence of the very transitions under discussion. Reference to table 13 shows indeed that if deactivating collisions are unimportant the resultant degree of vibrational excitation is actually rather too high. Though this might suggest that a satisfactory theory could be developed quantitative considerations reveal serious discrepancies. As is well known the transfer of vibrational energy does not normally occur at all rapidly. None the less ions can be deactivated readily by collision with their parent neutral molecules since the energy transfer can take place indirectly through charge exchange,

$$N_2^+(X\,^2\Sigma_g^+, v \neq 0) + N_2(X\,^1\Sigma_g^+, v = 0) \rightarrow N_2(X\,^1\Sigma_g^+, v \neq 0) + N_2^+(X\,^2\Sigma_g^+, v = 0). \quad (22)$$

Kallmann & Rosen (1929) have investigated charge exchange in nitrogen. They found the cross-section associated with the process to be $6 \cdot 3 \times 10^{-15}$ cm.^{2}. This value cannot be adopted without reserve as in the laboratory experiments the conditions were not identical with those in the upper atmosphere (the ions being relatively energetic and perhaps having a different distribution among the vibrational levels). If a cross-section of order gas kinetic is taken (and it appears improbable from Kallmann & Rosen's work that the true cross-section can be much less) and if the N_2 concentration at the appropriate altitude is assumed to be 10^{10}/cm.^{3}* then it can be seen that the number of deactivating collisions must be some 10^{-1}/ion/sec. The effect of these collisions on the relative population rates to N_2^+, $B\,^2\Sigma_u^+$ can be calculated from an equation similar to (18) but with certain additional terms (the form of which is obvious). Details need not be given here. The essential result is that the observed development of the band system can only be reproduced if the electron excitation rate is some 10 times the deactivation rate, that is only if each ion is excited on an average once a second. But the electron excitation coefficient cannot be greater than about 10^{-7} cm.^{3}/sec. Hence the required concentration of active electrons is some 10^7/cm.^{3}: and the concentration of passive electrons must be several orders greater† (cf. appendix). Such concentrations are of course impossible to accept. They would lead for example to an absurdly large rate of photon emission: and other difficulties occur. Finally, it may be mentioned that reasons can be given for supposing that direct excitation of nitrogen ions is in any case unlikely to be important except perhaps at extreme altitudes (cf. appendix).

So far in discussing the intensity distribution it has been assumed that the excitation arises through electron impact. If heavy particles such as protons are responsible (a possibility considered by Rayleigh 1922) the whole problem is altered, as in these circumstances the relative rates of population to the various vibrational levels are not the same as given in the tables used. Thus Smyth & Arnott (1930) have demonstrated qualitatively that there is a marked tendency for heavy particle excitation to favour transitions involving an increase in vibrational quantum number. This may conceivably provide the explanation of the high development of the negative band system in the low latitude aurorae. It is not, however, at present apparent, why in these in particular, excitation by heavy particles should be so important.

---

* This figure is uncertain because of the lack of knowledge regarding the degree of dissociation of nitrogen.

† Those electrons with sufficient energy to cause excitations or ionizations are classed as active: the remainder are classed as passive.

To summarize, Barbier's measurements on low latitude aurorae lead to very interesting conclusions. Seemingly the vibrational temperature of the molecules must be extremely high or else heavy particles must give a considerable contribution to the excitation. The first of these alternatives is rather the less probable, encountering as it does very grave difficulties. Until confirmatory work has been carried out it is not thought justifiable to attempt to elaborate the theory further and no final decision can be reached at present.

3·2·2. It has been known for some time that the intensity distribution of the negative system is different in low and high latitude aurorae. Thus Rayleigh (1922), in an investigation of the blue group ($v''-v' = 2$), found that while at Lerwick (Shetland) only the (0, 2) band λ4709 appeared on his photographs, at Terling (S. England) the (2, 4) band λ4600 was equal in intensity to it and the (1, 3) band λ4652 was actually stronger.

The only extensive intensity measurements on high latitude aurorae are those by Vegard and his associates in Norway. Vegard & Kvifte (1945) have published data on what they term to be a typical intensity distribution which was obtained from spectrograms taken with very long exposures arranged so that an average was obtained over a great variety of auroral forms. The altitude of the main luminescence was about 100 km., but there were also contributions from above this level. Table 17 gives the relative intensities (corrected at least approximately for atmospheric extinction) and table 18 the p's derived from them by the procedure already described.

Superficially the relative intensities seem to have the customary general trend—the (0, 0) band is, for example, the most prominent as usual, and in the main the bands present are those expected to be strong and the bands absent are those expected to be weak. However, quantitative examination shows that in fact the distribution is most abnormal. This is most apparent from the values of the p's,* which are very different from those calculated in the present paper (cf. table 4) or deduced from Herzberg's laboratory measurements (cf. table 5). As the agreement between these last is satisfactory and as Barbier's low latitude auroral data are in fair accord with them it is scarcely possible that it is they that are incorrect. Further, even if no such comparison data were available the p's derived from the high latitude aurorae would still be subject to grave suspicion as they do not satisfy the consistency tests† given in § 2·2. It can be seen for instance that $\sum_{v'} p(v', 0)$ does not approximate to unity as is required: and there are other similar failures. There is little doubt therefore that some disturbing factor has entered. What this may be is obscure.

Comparison between table 17 and table 6 shows that in any progression the auroral bands of shorter wave-length tend to be unduly weak compared with those of longer

---

* It need hardly be mentioned again that these are dependent only on the emission transition N_2^+, $B\,^2\Sigma_u^+ \rightarrow X\,^2\Sigma_g^+$—they are independent of the excitation mechanism.

† The assumptions on which these tests are based must, of course, be borne in mind.

TABLE 17. INTENSITY DISTRIBUTION OF NEGATIVE SYSTEM AS OBSERVED IN HIGH LATITUDE AURORAE[*]

(0 indicates band not recorded; – indicates the band outside spectral region studied)

v' \ v''	0	1	2	3	4	5
0	47·4	24·4	7·8	0	0	0
1	1·6	2·2	5·9	4·6	0	0
2	0	1·6	0	2·0	3·4	0
3	–	0	0	0	0	2·0

TABLE 18. p'S FOR N_2^+, $B\,{}^2\Sigma_u^+ - X\,{}^2\Sigma_g^+$ TRANSITION (HIGH AURORAE)

(In some cases it is only possible to say whether the p value is large (l) or small (s))

v' \ v''	0	1	2	3	4
0	0·48	0·35	0·17	s	s
1	0·05	0·10	0·39	0·45	s
2	s	0·11	s	0·26	0·63

wave-length.[†] This is the direction of the discrepancy that would be caused, if atmospheric extinction were grossly underestimated. But the effect is too great for such an explanation to be plausible. Thus the correction Vegard (1932) applies to the (0, 0) band λ3914 relative to that to the (0, 1) band λ4278 is only 1·2; and to bring the observed intensities into conformity with those in the present paper a further factor of about 1·9 is required. It might be suggested that self-absorption might be the process responsible. Since the relative populations of the various vibrational levels of the nitrogen ions must decrease with increasing vibrational quantum number a change of the correct sense in the intensity distribution is certainly produced in this way. However, an elementary calculation shows that quantitatively the effect is unimportant: further, the band profiles published by Vegard & Tønsberg (1935) give no indication of the phenomena. Suspicion is cast on the reliability of the auroral measurements by the fact that Vegard (1940) claims that the ratio of the intensities of bands *arising from the same vibrational level* is not a constant. This apparent effect is illustrated in table 19 which gives the average results of measurements made at Tromsø and Oslo on the intensities of the first three members of the $v' = 0$ progression. Both sets of figures differ appreciably from the corresponding set in table 17 and considerably from each other: thus, as will be noted, the intensity distribution at Tromsø exhibits a more rapid fall-off than does that at Oslo. That the difference is real is difficult to credit: it certainly cannot be associated with the actual emission in the aurorae itself: and as has been remarked the effects of atmospheric extinction and self-absorption are too small. Both quartz and glass spectrographs were used and yielded much the same type of result (cf. Vegard & Tønsberg

---

[*] The identification of the (5, 7) band with an emission of intensity 1·6 at λ4487 is also proposed but this must be treated with reserve. Isolated bands are always open to suspicion.

[†] The effect is strikingly illustrated if the ratios of corresponding intensities are tabulated.

1940). Nevertheless the possibility that instrumental or calibration error enters in some way must be raised. Some support is lent to it by the fact that Rypdal & Vegard (1940), employing a somewhat similar technique to that in the observational work, obtained from a laboratory source an intensity distribution resembling that given for the aurorae but quite different from that found by all other workers. The individual measurements too showed inconsistencies: for example, the ratio of the intensity of the (0, 0) band λ 3914 to that of the (0, 1) band λ 4278 ranged between 0·93 and 2·4.

TABLE 19. INTENSITY DISTRIBUTION OF THE $v' = 0$ PROGRESSION OF THE NEGATIVE SYSTEM AS OBSERVED AT DIFFERENT HIGH LATITUDE AURORAL STATIONS*

station	(0, 0) λ 3914	(0, 1) λ 4278	(0, 2) λ 4709
Tromsø	47·4	17·5	4·8
Oslo	47·4	29·5	12·1

The uncertainties mentioned make it impossible at present to discuss the excitation mechanism satisfactorily. However, a preliminary survey is at least possible. To minimize possible errors in the data bands of constant $(v' - v'')$ (and thus having wave-lengths not greatly different) were considered together: for each such group the relative rates of population of the various vibrational levels of N_2^+, $B\,^2\Sigma_u^+$ were computed from the intensities by the use of table 6; the values obtained were then averaged. Table 20 contains the final results of the analysis. Information is scanty for the level $v' = 3$ as only a single band, (3, 5) λ 4554, arising from it has been recorded. It is desirable that the (3, 2) band λ 3549 which is the strongest in the progression, be sought.

TABLE 20. RELATIVE RATES OF POPULATION $g(v')$ OF THE VARIOUS VIBRATIONAL LEVELS OF N_2^+, $B\,^2\Sigma_u^+$ (HIGH LATITUDE AURORAE)

v'	0	1	2	3
$g(v')$	1·00	0·25	0·12	(0·05)

The difference between the figures in table 20 and the corresponding ones in table 16 is very pronounced. It illustrates clearly the fact, already mentioned, that the negative system is much less fully developed in high latitude aurorae than in low latitude aurorae. However, even in high latitude aurorae the degree of the development appears to be considerable. If the excitation arises simply through collisions between electrons and nitrogen molecules in their ground state a vibrational temperature of some 2000° K is required to reproduce it approximately (cf. table 11). This is, of course, far below the vibrational temperature of 8000° K that had to be invoked in low latitude aurorae when the same hypothesis regarding the process of excitation was adopted. But the comparatively great particle concentration in the region near the 100 km. level makes it even more difficult to accept a vibrational temperature much above the translational temperature: and this latter is known to be about 500° K (or less). It would seem, therefore, that the alternative to the enhanced vibrational temperature theory, that is, heavy particle excitation, must be considered very seriously. The contribution from this to the total excitation need not be as great as in low latitude aurorae.

---

* The units were adjusted to make the intensity of the (0, 0) band λ 3914 the same as given in table 17. The separate results giving the means quoted show a scatter sufficiently wide to include some that agree with prediction. Thus, on one occasion Vegard obtained the following distribution: (0, 0) λ 3914—47·4, (0, 1) λ 4278—14·3, (0, 2) λ 4709 not recorded so presumably very weak. This is very close to that in table 6. Unfortunately no other bands on the plate were measured.

In view of the interest attached to the interpretation of the results it is desirable that the typical auroral intensity distribution be further investigated so that the apparent anomalies to which attention was drawn at the beginning of the present discussion may be understood and eliminated. Until this is done the differences between tables 11 and 20 can only be regarded as suggestive.

The question naturally arises as to whether or not the intensity distribution of the negative system in high latitude aurorae depends on the altitude. To answer this it is necessary, as stressed by Nicolet,[*] to have data obtained from short exposures with the spectrograph pointed in a fixed direction, and at a particular aurora. The Norwegian workers have published very few results got under the conditions specified. Four sets of suitable measurements have, however, been given by Vegard & Kvifte (1945)—two from a region close to the base of aurorae (low altitude) and two from near the convergence point of coronal forms (high altitude). There was no marked difference in the intensity distribution.[†] This, if confirmed, is a very important result which may have a considerable bearing on auroral theory (in particular in connexion with the heavy particle content of the incident streams). It is contrary to the expectation that the negative system in high latitude aurorae of great altitude might show the same extraordinary development as in low latitude aurorae (which are also of great altitude).

3·2·3. The relative intensity of the negative bands as emitted by high sunlit auroral rays has been investigated by Størmer (1939). He confined his attention mainly to the intensity distribution within the blue-violet ($v''-v'=1$) and blue ($v''-v'=2$) groups; since the total wave-length range in either group is small he was able to obtain the relative intensities directly from his photometric record without introducing any corrections: he made no attempt to compare the two groups and merely expressed his results in arbitrary units chosen so that 100 represented the intensity of the leading members $\lambda 4278$ and $\lambda 4709$.

The excitation mechanism operative is certainly the resonance process,

$$\mathrm{N}_2^+(X\,{}^2\Sigma_g^+)+h\nu\rightarrow\mathrm{N}_2^+(B\,{}^2\Sigma_u^+). \tag{23}$$

As the average time between successive absorptions is rather less than 10 sec. (Bates 1949), collisions are clearly too infrequent at the great altitudes concerned (between 400 and 650 km.) to prevent the equilibrium degree of vibrational excitation of the ions being established.* Hence from tables 13 and 6 the intensity distribution to be expected can be derived.

---

[*] Private communication.

[†] As a reservation to this statement it may be mentioned that for the low altitude aurorae intensities were not given for either the (2, 3) band $\lambda 4199$ or for the (2, 4) band $\lambda 4600$ which might mean that they were too feeble to measure. However, for the high altitude aurorae their intensities relative to those of the other bands was the same for the typical distribution.

Table 21 shows the relative intensities as given by Størmer and as predicted in the manner just indicated. In view of the approximate nature of the intensity measurements the general agreement shown is quite good. It is very satisfactory that the theory can account so well for the unusually high intensities of the later members of the two groups.† The observed development actually appears to be somewhat fuller than the calculated, but it is not possible to state whether or not the difference is of real significance.

TABLE 21. RELATIVE INTENSITIES IN HIGH SUNLIT AURORAL RAYS

	blue-violet group			blue group	
band	measured intensity	predicted intensity	band	measured intensity	predicted intensity
(0, 1) λ4278	100	100	(0, 2) λ4709	100	100
(1, 2) λ4236	59	51	(1, 3) λ4652	97	74
(2, 3) λ4199	35	24	(2, 4) λ4600	85	78

3·2·4. The negative band system has been found to be present in the spectrum of the twilight sky. There is now little doubt (cf. Bates (1949)) that the excitation process involved is the same as that just discussed. However, Dufay & Dufay (1947) have shown that the altitude of the atmospheric region responsible is only about 100 km., and here the molecular concentration is so great that the relative populations in the various vibrational levels of the ions are mainly controlled by collisions and not by absorption and emission of resonance radiation. Hence, in calculating the intensity distribution, table 12 must be used in place of table 13. As the vibrational temperature is presumably low the development of the system would not be expected to be as full as in the high sunlit auroral rays: but it would be expected to be more extensive than if the excitation process were by electronic bombardment. Though Dufay & Dufay do in fact remark that the development is considerable they do not give any actual intensity measurements. It is very desirable that these should be made.

4. THE SECOND POSITIVE SYSTEM

4·1. *The basic data*

Table 22 shows the wave-lengths of the second positive system $C\,^3\Pi_u$–$B\,^3\Pi_g$ as determined by Pankhurst & Gaydon.‡ The calculated p's are given in table 23. Estimated intensities have been published by Birge (1926), Smyth & Arnott (1930) and others, but reliable quantitative measurements are somewhat lacking. The most extensive series is that due to Tawde (1934) who investigated the intensity distribution from four types of source and thereby derived the p's associated with the transition. Unfortunately, the values he obtained from the various sources differed very greatly so that they cannot be used with much confidence. Nevertheless it is perhaps worth quoting the averaged p's and this is done in table 24. As can readily be verified they do not satisfy the consistency test of § 2·2. However, comparison with table 23 suggests that they are not grossly in error. They may therefore be used at least as an approximate guide for those bands for which no other information is available.

---

* It may be mentioned that collision deactivation is less important for photo excitation than for electron excitation (*v. supra*) as in the equilibrium distribution (neglecting collisions) the relative population of the levels with $v \neq 0$ is much less.

† The extent of this enhancement can best be realized by comparison of the figures in table 21 with the corresponding figures in table 9 which gives the intensity distribution for what may be regarded as the standard excitation mechanism (collision between electrons and nitrogen molecules at a low vibrational temperature).

‡ Quoted in Pearse & Gaydon (1941).

TABLE 22. WAVE-LENGTHS OF BANDS OF SECOND POSITIVE SYSTEM (Å)

v' \ v''	0	1	2	3	4	5	6	7	8	9
0	3371	3577	3805	4059	4344	4667				
1	3159	3339	3537	3755	3998	4270	4574	4917		
2	2977	3136	3309	3500	3710	3943	4201	4490	4815	
3	2820	2962	3117	3285	3469	3672	3895	4142	4417	4723
4		2814	2953	3104	3268	3446	3642	3858	4095	4355

TABLE 23. p'S FOR N_2, $C\,^3\Pi_u$–$B\,^3\Pi_g$ TRANSITION (THEORY)

v' \ v''	0	1	2	3
0	0·48	0·32	0·14	$0{\cdot}04_5$
1	0·39	$0{\cdot}03_8$	0·22	0·19
2	0·12	0·37	$0{\cdot}00_9$	$0{\cdot}09_4$

TABLE 24. p'S FOR N_2, $C\,^3\Pi_u$–$B\,^3\Pi_g$ TRANSITION (AVERAGED LABORATORY MEASUREMENTS)

(— indicates outside spectral range investigated)

v' \ v''	0	1	2	3	4	5	6	7	8	9
0	0·40	0·36	0·17	$0{\cdot}05_8$	$0{\cdot}00_8$	$0{\cdot}00_3$	$0{\cdot}00_0$			
1	0·49	$0{\cdot}02_6$	0·16	0·20	0·10	$0{\cdot}02_7$	$0{\cdot}00_6$	$0{\cdot}00_1$		
2	0·29	0·33	$0{\cdot}03_0$	$0{\cdot}05_9$	0·14	0·10	$0{\cdot}03_8$	$0{\cdot}01_1$	$0{\cdot}00_4$	
3	0·16	0·44	0·24	$0{\cdot}07_4$	0·14	$0{\cdot}06_4$	$0{\cdot}07_7$	$0{\cdot}03_8$	$0{\cdot}01_6$	$0{\cdot}00_7$
4	—	—	0·45	0·23	0·12	$0{\cdot}01_0$	$0{\cdot}05_2$	$0{\cdot}05_9$	$0{\cdot}05_3$	$0{\cdot}02_3$

Only two processes for populating the upper $C\,^3\Pi_u$ state were considered of sufficient interest to investigate. These are electronic excitation of the neutral molecule (in its ground state),

$$N_2(X\,^1\Sigma_g^+) + e \rightarrow N_2(C\,^3\Pi_u) + e, \tag{24}$$

the calculated p's for which are given in table 25, and electronic recombination with the ionized molecule (in its ground state),

$$N_2^+(X\,^2\Sigma_g^+) + e \rightarrow N_2(C\,^3\Pi_u) + h\nu, \tag{25}$$

the calculated p's for which are given in table 26. As the transitions are of the perpendicular type the figures should not be regarded as very accurate.

A check on some of the work is provided by the experiments of Langstroth (1935) who measured the intensity distribution resulting from process (24). Owing to the characteristic rapid initial rise of the excitation function associated with an optically forbidden transition, the ratios of the rates of population of the various vibrational levels of the upper state show some dependence on the energy of the bombarding electrons. The effect can most simply be expressed by introducing within the integral in equation (12) a multiplying factor $\mu(v', v'', \epsilon)$. Table 27 gives the values of the factor obtained by Langstroth. As is apparent, they tend to a constant (which can be taken to be unity) in the region a little above the excitation potential. The observed relative intensities here are given in table 28 and the predicted relative intensities (as derived from tables 22, 25 and the mean of tables 23 and 24) in table 29. It can be seen that the agreement between the two, though in general fair, is rather poor for the weaker bands: there is moreover a tendency for the predicted development to be less extensive than the observed. The fact (already noted) that the excitation transition is of the perpendicular type may account for some of the differences. As little further can be done theoretically to improve the position, additional laboratory measurements should be performed. At present, to minimize the degree of uncertainty, it is probably best to average the data.

TABLE 25. p's FOR N_2, $C\,^3\Pi_u$–$X\,^1\Sigma_g^+$ TRANSITION (THEORY)

v' \ v''	0	1	2
0	0·52	0·36	0·10
1	0·31	$0{\cdot}05_0$	0·35
2	0·12	0·23	$0{\cdot}01_9$

TABLE 26. p's FOR N_2^+, $X\,^2\Sigma_g^+$–N_2, $C\,^3\Pi_u$ TRANSITION (THEORY)

v' \ v''	0	1	2
0	0·80	0·16	$0{\cdot}02_7$
1	0·18	0·47	0·26
2	$0{\cdot}01_3$	0·32	0·22

As for the negative system it is useful to know the intensity distribution that arises if the population rates to the various vibrational levels of the excited state are equal. Table 30 was therefore compiled. It was obtained by averaging the results from table 22 and the mean of tables 23 and 24 and from table 28. In employing it, the possible inaccuracies of the data on which it is based must be remembered—the figures at the extreme right (which are based solely on Tawde's measurements) are especially open to suspicion.

The different tables can, of course, be combined together to yield results for a number of excitation conditions other than those already discussed. Table 31 is of some importance. It gives the relative population rates for process (24) with the parent molecules not in their zeroth vibrational level (as previously assumed) but instead distributed among the different vibrational levels, their effective temperature ranging from 500 to 8000° K. An interesting feature is that taking the population rate to the zeroth vibrational level to be unity the relative population rate to the first vibrational level initially *decreases* with rise of temperature. This

TABLE 27

electron energy ϵ (eV)	multiplying factor $\mu(v', v'', \epsilon)$		
	$v' = 0$	$v' = 1$	$v' = 2$
14	1·00	0·68	0·52
15	1·00	0·72	0·60
16	1·00	0·85	0·84
$\geqslant 30$	1·00	1·00	1·00

TABLE 28. OBSERVED INTENSITY DISTRIBUTION OF SECOND POSITIVE BANDS AS EXCITED BY THE COLLISION OF ELECTRONS WITH NEUTRAL NITROGEN MOLECULES IN THEIR GROUND ELECTRONIC STATE AND ZEROTH VIBRATIONAL LEVEL*

(— indicates band outside spectral region studied)

v' \ v''	0	1	2	3	4	5
0	19	16	6·2	1·9	—	—
1	—	$0{\cdot}4_8$	5·6	5·2	2·7	—
2	—	—	$0{\cdot}3_6$	$0{\cdot}7_6$	1·7	1·3

TABLE 29. PREDICTED INTENSITY DISTRIBUTION OF SECOND POSITIVE BANDS AS EXCITED BY THE COLLISION OF ELECTRONS WITH NEUTRAL NITROGEN MOLECULES IN THEIR GROUND ELECTRONIC STATE AND ZEROTH VIBRATIONAL LEVEL

v' \ v''	0	1	2	3	4	5	6	7
0	23	14	5·1	1·3	0·2	0·0		
1	17	1·0	4·6	3·7	1·4	0·3	0·0	
2	3·3	4·7	0·2	0·7	0·9	0·5	0·1	0·0
3	0·9	2·1	0·9	0·2	0·3	0·1	0·1	0·0

TABLE 30. PREDICTED INTENSITY DISTRIBUTION OF SECOND POSITIVE BANDS ASSUMING EQUAL RATES OF POPULATION OF THE VIBRATIONAL LEVELS OF THE EXCITED STATE

v' \ v''	0	1	2	3	4	5	6	7	8
0	41	29	11	3·1	0·3	0·1			
1	54	2·2	14	12	5·6	0·9	0·2	$0{\cdot}0_3$	
2	28	39	1·7	4·6	7·7	5·1	1·2	0·3	0·1
3	18	41	18	4·6	7·0	2·6	2·4	0·9	0·3

* The units are adjusted so that the total photon emission in those bands for which the intensities are available is equal to that in the same bands as given by table 29.

abnormal effect is caused by the smallness of $p(1, 1)$ relative to $p(0, 1)$ (cf. table 25). The usual type of variation is exhibited by the relative population rate to the second vibrational level. It will be noted that the influence of temperature is much less marked than in the case of the negative system. The accuracy of the figures is not high (*v. supra*).

The measurements at present available on the emission of the second positive system from the upper atmosphere are not sufficiently extensive to justify publishing further auxiliary tables here. If required, these can readily be derived from the data that have been collected.

TABLE 31. RELATIVE RATES OF POPULATION OF THE VARIOUS VIBRATIONAL LEVELS OF N_2, $C\,^3\Pi_u$ BY ELECTRON IMPACT WITH N_2, $X\,^1\Sigma_g^+$

effective vibrational temperature (° K) ...	500	1000	2000	4000	6000	8000
$v' = 0$	1·00	1·00	1·00	1·00	1·00	1·00
1	0·60	0·58	0·56	0·60	0·66	0·71
2	0·23	0·24	0·28	0·35	0·43	0·48

4·2. *Discussion*

The spectral location of the second positive system is not favourable for observational work, many of the stronger bands being in the ultra-violet region. Consequently extensive measurements have not been made.

4·2·1. As far as low latitude aurorae are concerned the literature appears only to contain a few isolated remarks on identifications. Dufay & Tcheng Mao-Lin (1942) found evidence for the presence of the (0, 2) band λ3805. They pointed out that, as this is situated near the short wave-length limit of their instrument, its intensity must be high for it to have appeared at all. No other bands of the system were detected on their spectra. Reference to the tables in the previous section shows that this is consistent with what would be expected from any normal excitation mechanism. Barbier (1947) extended the spectral range beyond that covered by Dufay & Tcheng Mao-Lin and tentatively suggested the presence of not only the (0, 2) band λ3805 but also of the (0, 0) band λ3371. He does not make any mention of the (0, 1) band λ3577 whose intensity must be between that of the other two, but conceivably it might have been obscured by the (1, 0) band of the negative system λ3582. In any future work it is desirable that a careful search be made for the individual bands that are expected to be the most intense. With the plausible assumption that the excitation is similar to that of the negative system, it can be seen from the data collected in § 4·1 that the following are likely to be the most prominent of the bands of wave-length above 3000 Å: (0, 0) λ3371, (0, 1) λ3577, (0, 2) λ3805, (1, 0) λ3159, (1, 2) λ3537, (1, 3) λ3755 and (2, 1) λ3136. The detection of radiations at these wave-lengths would fully establish the presence of the second positive system in low latitude aurorae. Information on the intensities should provide a useful guide to the excitation conditions if used in conjunction with tables 23 to 31; it will be particularly interesting to compare any conclusions reached in this way with those deduced from the intensity distribution of the negative system. Finally it is perhaps worth drawing attention to the fact that the (3, 4) band λ3469 is comparatively weak and is only likely to be observable if a number of other members of the system, including those mentioned above, are strongly emitted (which they do not normally appear to be). It is therefore improbable that the radiation at about* λ3471Å found by Götz (1941) is due to this band. The correct identification is doubtless with the forbidden doublet of atomic nitrogen $^2P-^4S$, λ3466 whose presence in low latitude aurorae was first observed by Slipher & Sommer (1929) (cf. discussion by Dufay & Tcheng Mao-Lin 1942 and Dufay 1943).

TABLE 32. INTENSITY DISTRIBUTION OF SECOND POSITIVE SYSTEM AS OBSERVED IN HIGH LATITUDE AURORAE†

(0 indicates band not recorded; *f* indicates feeble intensity; – indicates band outside spectral region studied)

v' \ v''	0	1	2	3	4	5	6	7
0	9·0	9·8	4·9	3·4	3·0	0	0	0
1	5·3	1·2	4·9	4·2	3·7	0	0	0
2	–	3·6	0	2·2	2·4	2·2	0	0
3	–	–	*f*	1·8	3·0	*f*	0	1·4

4·2·2. Vegard & Kvifte (1945) have published a typical intensity distribution for the second positive system in high latitude aurorae. This is reproduced in table 32: the units are the same as in table 17. There is some general correlation between the observed and predicted (table 29) variation of intensity along any v'' progression. Thus the minima at the (1, 1) band λ3339 and the (2, 2) band λ3309 occur in both. The agreement is not, however, detailed. As in the negative system the bands to the left tend to appear enfeebled relative to those to the right.‡ In view of the discussion in § 3·2·2 this will not be elaborated upon here. It may, however, be mentioned that even allowing for the anomalous trend the (0, 4) band λ4344 and the (3, 7) band λ4142 seem much too intense—possibly some other emission contributes.

The nature of the data available makes it impossible to derive accurate relative population rates to the various vibrational levels of N_2 $C\,^3\Pi_u$. Nevertheless some suggestive results can be obtained by combining the information in tables 29 and 32 and averaging over groups of bands of about the same wave-length. These are given in table 33. The slowness of the decrease of the relative population rate with increasing v' is very striking. In any comparison with table 20 it should be remembered that because of fundamental properties of the nitrogen molecule the negative system is the more compact naturally.

Mitra (1946) and Ghosh (1946) have suggested that the emission of the second positive system is due to the formation of nitrogen molecules in the $C\,^3\Pi_u$ state by recombination between electrons and nitrogen ions in the $X\,^2\Sigma_g^+$ state. Bates, Massey & Pearse (1947) have criticized this proposal on the grounds that the photon yield is quite inadequate. Further evidence against it is provided by considering the degree of development of the system that would ensue. This can be calculated with the aid of table 26 (cf. § 2·2). Taking a vibrational temperature of 500° K or less, then with $g(0)$ unity, $g(1)$ is only 0·17 and $g(2)$ is only 0·024. These values are far below those deduced from the auroral data. The discrepancy cannot be removed by any acceptable increase in the vibrational temperature.

---

* The wave-length measurement is not precise as the dispersion of the spectrograph used was very low.

† The presence of a few bands additional to those in the table has been suggested. Thus the (7, 8) and (9, 7) bands have been identified with weak emissions at λ3484 and λ3169 respectively. The proposals are not at all plausible and will be ignored.

‡ From the data in the two systems a crude (intensity correction)-(wave-length) curve can be derived but the scatter is too great to make it of much value.

TABLE 33. RELATIVE RATES OF POPULATION $g(v')$ OF THE VARIOUS VIBRATIONAL LEVELS OF N_2, $C\,^3\Pi_u$ (HIGH LATITUDE AURORAE)

v'	0	1	2	3
$g(v')$	1·00	$0{\cdot}9_1$	$0{\cdot}8_0$	$(0{\cdot}6_8)$

Electron impact with nitrogen molecules in their ground state $X\,^1\Sigma_g^+$ is less unsatisfactory than the recombination process, but it also fails to provide the development apparently required. Table 31 gives the predicted relative population rates. As can be seen, they are much smaller than the corresponding values in table 33. Their peculiar dependence on the vibrational temperature makes it even more probable than in the case of the negative system that an enhancement of this by some overlooked effect cannot be invoked to explain the observed intensity distribution. But unfortunately, owing to the excitation transition being of the perpendicular type, the uncertainty in the calculated p's makes the reality of the difference open to doubt. Further direct laboratory measurements are essential.

Nevertheless, it would once again seem reasonable to suppose that heavy auroral particles give a considerable contribution to the excitation. In this case, however, there is the alternative possibility that ionic recombination

$$N_2^+(X\,^2\Sigma_g^+)+O^-(^2P)\rightarrow N_2(C\,^3\Pi_u)+O(^3P,\,(\text{or}\,^1D)) \tag{26}$$

is responsible: but it should be noted that simple charge transfer is not sufficient for this process—the $X\,^2\Sigma_g^+$ configuration $(K, K, \sigma_g^2, \sigma_u^2, \pi_u^4, \sigma_g)$ and the $C\,^3\Pi_u$ configuration $(K, K, \sigma_g^2, \sigma_u, \pi_u^4, \sigma_g^2, \pi_g)$ are such that in addition an internal rearrangement is required. It would therefore seem probable that the effective collision coefficient is not large. Attention may be drawn to the fact that the transition from the $X\,^1\Sigma_g^+$ state to the $C\,^3\Pi_u$ state involves a change in multiplicity. Account must be taken of this in assessing the plausibility of heavy particle excitation. The transition could not be produced by simple proton impact. But it could be produced by protons that had captured an electron (or by any other complex system) as electron exchange could then occur. The cross-section associated with such a process requires investigation.

4·2·3. Finally, before concluding the discussion, it is perhaps worth making a brief reference to the night sky emission. Most workers agree that the second positive system is normally absent. But the complexity of the spectrum is such that it is difficult to be certain. Dufay & Déjardin (1946) list 12 wave-length coincidents. By using table 30, which gives the relative intensities of the bands in any v'' progression, the plausibility of the identifications can be judged. Details of the comparison need not be given here: it is sufficient to state that many anomalies appear. This gives support to the view of Dufay & Déjardin that the bands in question are best attributed to other systems.

5. THE FIRST POSITIVE SYSTEM

5·1. *The basic data*

Only the infra-red bands of the first positive system $B\,^3\Pi_g$–$A\,^3\Sigma_u^+$ will be considered here. Table 34 gives the wave-lengths of some of the more important ones as determined by Poetker (1927). Table 35 gives their calculated p's. Now in addition to his wave-length measurements Poetker also obtained values for the relative intensities. While these are not of high accuracy* approximate p's can be estimated from them. The values obtained are shown in table 36. As can be seen, the agreement with table 35 is not detailed but at least there is confirmation of its main features which is as much as could be expected in view of the character of the

TABLE 34. WAVE-LENGTH OF BANDS OF FIRST POSITIVE SYSTEM (Å)†

v' \ v''	0	1	2
0	10,420	12,220	(14,790)
1	8,860	(10,140)	11,830
2	7,750	8,670	9,880

TABLE 35. p'S FOR N_2, $B\,^3\Pi_g$–$A\,^3\Sigma_u^+$ TRANSITION (THEORY)

v' \ v''	0	1	2
0	0·32	0·32	0·20
1	0·41	$0{\cdot}00_7$	$0{\cdot}08_6$
2	0·21	0·19	0·13

TABLE 36. p'S FOR N_2, $B\,^3\Pi_g$–$A\,^3\Sigma_u^+$ TRANSITION (LABORATORY MEASUREMENTS)

v' \ v''	0	1	2
0	0·35	0·35	0·14
1	0·27	0·00	0·25
2	0·15	0·32	0·07

---

* The intensities, which range between 1 and 10 (in arbitrary units) are only given to the nearest whole number.

† These wave-lengths correspond to the band maxima under laboratory conditions: somewhat different values may be found under upper atmospheric conditions.

experimental results and of the fact that the transition is of the perpendicular type. In applications it is probably justifiable to assume that the calculated p's are fairly reliable.

Excitation processes were not investigated and it will only be remarked that the difference in the equilibrium inter-nuclear separation of $N_2 X\,{}^1\Sigma_g^+$ and $N_2 B\,{}^3\Pi_g$ is comparatively large (being in fact $0{\cdot}115 \times 10^{-8}$ cm.) so that the bombardment of nitrogen molecules in their ground state by electrons leads preferentially to the population of the higher vibrational levels.

5·2. *Discussion*

One of the most remarkable features of the night sky spectrum is the great infra-red emission near λ 10,440 discovered by Herman, Herman & Gauzit (1942). Stebbing, Whitford & Swings (1945), who independently observed the radiation, made the suggestion that it should be identified as the (0, 0) band of the first positive system, λ 10,420. Now the (1, 0) band λ 8860 does not appear to be present with any appreciable intensity. This might at first seem peculiar, but Stebbing, Whitford & Swings put forward the explanation that a selective excitation mechanism is operative. They suggested that the process responsible is simply

$$N(^4S) + N(^4S) + N_2(X\,{}^1\Sigma_g^+) \rightarrow N_2(B\,{}^3\Pi_g) + N_2(X\,{}^1\Sigma_g^+). \tag{27}$$

If the dissociation energy of nitrogen is 7·383 eV* the excited molecules $N_2(B\,{}^3\Pi_g)$ can only be in their zeroth vibrational level, thus satisfying the observational requirement.

While the identification mentioned is attractive further work is necessary before it can be regarded as established. In particular it is desirable that a search be made for the (0, 1) band near λ 12,220 and the (0, 2) band near λ 14,790. From the data in § 5·1 it can be seen that the intensities of these two bands should be about 0·53 and 0·15 respectively (the intensity of the (0, 0) band λ 10,420 being taken to be unity). The observation of radiations of the wave-length and intensities stated would provide conclusive evidence for the emission from the night sky of the $v' = 0$ progression of the first positive system. In the meanwhile, the hypothesis of this emission forms a convenient basis for discussion.

The lower state of the first positive system is, of course, the upper state of the Vegard-Kaplan system $A\,{}^3\Sigma_u^+ \rightarrow X\,{}^1\Sigma_g^+$. This latter system should therefore also be emitted from the night sky. While it has indeed been found to be present the intensity distribution is not what would arise from the cascade effect. This is illustrated by table 37 which gives (*a*) the predicted relative rates of populating the various vibrational levels of N_2, $A\,{}^3\Sigma_u^+$ by the $v' = 0$ progression of the first positive system and for comparison (*b*) the sum of the intensities (in arbitrary units) of the bands originating from each as observed by Cabannes & Dufay (1946). The figures demonstrate clearly that cascading is not the dominant mechanism responsible for bringing nitrogen molecules to the $A\,{}^3\Sigma_u^+$ state—a view expressed earlier by Bates (1947):

---

* Gaydon (1947) supports a larger value, 9·764 eV.

some other process, at least 10 times as effective, must be operative. A number of possibilities suggest themselves (cf. Swings 1947), the most plausible being

$$\mathrm{N}(^4\mathrm{S})+\mathrm{N}(^4\mathrm{S})+X\rightarrow \mathrm{N}_2(A\,^3\Sigma_u^+)+X. \tag{28}$$

Enough energy is available for the excitation of the vibrational levels up to $v'=8$ (compare table 37): there is insufficient for the excitation of the vibrational level $v'=9$, the highest from which emission has been suggested to rise, but little importance need be attached to this.

TABLE 37

v' (of N_2, $A\,^3\Sigma_u^+$) ...	0	1	2	3	4	5	6	7	8	9
(*a*) predicted relative population rates (by cascading)	34	21	8	(2)	(1)	—	—	—	—	—
(*b*) sum of band intensities	4	5	15	12	15	5	5	3	1	1

The radiation at λ 10,440 is extremely strong compared with any other in the night sky spectrum. Stebbing, Whitford & Swings estimate it to be perhaps 100 times more intense than even λ 5577, the prominent green forbidden line of atomic oxygen ^1D–^1S; and to obtain the ratio of the photon emissions (including the emission from the unobserved part of the first positive system) it can readily be seen that a further factor of almost 4 must be introduced. Thus since according to Rayleigh (1930) the number of transitions needed for the green line of oxygen is 2×10^8/cm.^{2} column/sec. the number needed for the first positive system of nitrogen must be some 8×10^{10}/cm.^{2} column/sec. If the thickness of the luminous layer is taken as 10 km. the approximate rate of process (27) must therefore be 8×10^4/cm.^{3}/sec., and hence that of process (28) must be 8×10^5/cm.^{3}/sec.* From these figures it can be seen that about 8×10^{10} nitrogen atoms/cm.^{3} would be lost during the course of a night.† Now Stebbing, Whitford & Swings found that the infra-red emission was maintained throughout the hours of darkness: their photo-cell current only fell from 100 to 150 units just after sunset to 70 to 100 units just before sunrise. This indicates that the fraction of the free nitrogen atoms that disappear in a night is small, perhaps a fifth. Hence the peak concentration of atomic nitrogen must be of order 4×10^{11}/cm.^{3}.

It follows from the nature of process (27) and (28) that the altitude at which the peak is located must be quite low. As little is known of the factors that determine the efficiency of three body reactions one would hesitate to estimate the coefficient associated with process (27) since the close resonance that occurs might conceivably greatly influence it; but process (28) (which, as has been seen, is the more rapid) presents no unusual features, so that its coefficient can scarcely be greater than

---

* The relatively feeble intensity of the Vegard-Kaplan system makes it necessary to assume that only an extremely small fraction (perhaps about 1 in 10^4) of the metastable molecules radiate, the remainder being deactivated by collisions. This requires the combination of a low spontaneous transition probability and a high collision deactivation coefficient.

† The number given may be an underestimate as nitrogen atoms may be removed by processes other than (27) and (28). This possibility which adds to certain difficulties must be remembered in the subsequent discussion.

10^{-31} cm.^{6}/sec.* If this value is adopted, and if the atomic nitrogen concentration is taken to be 4×10^{11}/cm.^{3}, then to obtain the association rate of 8×10^5/cm.^{3}/sec. a third body concentration of 5×10^{13}/cm.^{3} is needed, and this occurs about the 100 km. level. The Vegard-Kaplan system should, of course, also be emitted from here. This conflicts with the measurements of Barbier (1947) who claims that the altitude of their source is from 350 to 900 km. However, height determinations are still subject to much uncertainty. Further, Cabannes & Dufay (1946) have shown that the rotational temperature of the metastable molecules is only about 230° K which favours a low altitude for the emission.

The atomic nitrogen concentration involved is very high. It is not easy to understand how it could be established against the loss processes acting (to overcome which atoms must be freed at a mean rate of order 10^{12}/cm.^{2} column/sec. or even more). Several mechanisms by which nitrogen molecules can be dissociated during the day have been suggested by Ta-You Wu (1944) and by Herzberg & Herzberg (1948) but they appear too slow, involving as they do either photo dissociation by solar radiation of very short wave-length (less than 500 Å) or predissociation through a forbidden electronic transition ($X\,{}^1\Sigma_g^+ \rightarrow a\,{}^1\Pi_g$). The inadequacy of such processes is best illustrated by noting that the required rate is not greatly less than that for the dissociation of molecular oxygen which is produced by the action of radiation of comparative long wave-length (the threshold being 1760 Å) of very high intensity and which takes place through an allowed transition ($X\,{}^3\Sigma_g^- \rightarrow B\,{}^3\Sigma_u^-$). It should perhaps be mentioned that even if the infra-red emission is ignored the presence of certain other bands in the night sky spectrum is most readily explained if it is assumed that the excitation energy is provided by the association of free nitrogen atoms (cf. Bates 1947). However, a much smaller rate of formation of these latter suffices.

The remarkable nature of the conclusions regarding the state of the upper atmospheric nitrogen that appear to follow from the hypothesis of Whitford, Stebbings & Swings serves to emphasize the importance of further observational work to prove or disprove the correctness of identification of the infra-red emission. It is clearly also desirable for a precise absolute intensity measurement to be made and for the intensity distribution in the Vegard-Kaplan system to be further investigated. In addition the temporal variations of the infra-red emission, and of the Vegard-Kaplan system should be studied to verify if they correspond as would be expected if they originate as postulated.

In conclusion I would like to thank Professor H. S. W. Massey, F.R.S., for his continued interest and encouragement.

---

* This is about the magnitude that is usually taken for the coefficient for the total three-body association rate. It can readily be seen that a much smaller value leads to serious difficulties. Now in the current theory of the night sky emission a number of lines and bands are presumed to originate through the association of free oxygen atoms. To explain the observed intensities the partial coefficients for such associations need only be a small fraction of the total coefficient (unless there is very effective collision deactivation, and this is improbable at least for O (^1S) in view of the observation of Dufay & Dufay (1948) that the forbidden green line is enhanced at twilight to the extent predicted by simple theory). It is not altogether satisfactory that it has to be assumed that all or almost all the nitrogen associations yield particular excitations—especially as process (27), which might have been exceptional, is not the only one that has to be invoked.

APPENDIX

The energy distribution of the auroral electrons

The problem of calculating exactly the energy distribution of the auroral electrons is one of some complexity. No attempt is made here at a detailed solution which would require among other things a fuller knowledge of certain reaction rates than is available. It may be useful, however, to give the results of an elementary exploratory investigation. Heavy particles will be ignored but this does not affect the essential argument.

As a crude model it is convenient to divide the auroral electrons into one of two main groups according to whether or not they have energy sufficient to excite or ionize the atmospheric constituents. Following the terminology used in the text the electrons in the high energy group will be referred to as active and those in the low energy group as passive: $n(eA)$ and $n(eP)$ denote their respective concentrations. Consideration of the equilibrium is sufficient to enable an estimate to be made of the ratio of these concentrations. Essentially the electrons in the active group are constantly being degraded to the passive group by loss of energy through excitation and ionization. The entry of electrons into the passive group must balance the disappearance of electrons from it due to recombination and attachment processes. This is expressed by the equation

$$\left.\begin{aligned}\zeta n(eA)\,n(X) &= \alpha_e n(eP)\,n(X^+) + \eta n(eP)\,n(X)\\ &= \alpha_e n(eP)\{n(eP)+n(eA)\}(1+\lambda)+\eta n(eP)\,n(X),\end{aligned}\right\} \quad (\mathrm{A}\,1)$$

where $n(X)$ and $n(X^+)$ are respectively the concentrations of neutral particles and positive ions, λ is the negative ion-electron ratio, ζ is the coefficient associated with the degradation processes, α_e is the electronic recombination coefficient, and η is the attachment coefficient. It need scarcely be emphasized that the situation that actually prevails has been greatly simplified. For example, the different atmospheric constituents are treated as one, and a number of processes such as collision detachment and ionic recombination are ignored.* In spite of this equation (1) probably serves as a useful guide. Before proceeding to use it values have to be assigned to the various coefficients.

Electrons in the higher part of the active group have to make many inelastic collisions (about 50 for those of energy 1 kV) before they themselves are finally degraded. This might suggest that ζ is much less than the inelastic collision coefficient. However, the electrons that are ejected in ionizing collisions (which form a considerable fraction of the total number of inelastic collisions) tend to be in the lower part of the active group and are degraded by only a few inelastic collisions. It is therefore reasonable to assume that the rate of passage of electrons from the active group to the passive group does not differ by a large factor from the inelastic collision rate: that is that ζ is only several times smaller than the inelastic collision coefficient. On this basis a figure of 10^{-7} cm.^{3}/sec. was adapted.† Reliable experimental or theoretical information on α_e not being available it seems best to employ the values given by ratio observers, 10^{-8} cm.^{3}/sec. just above 100 km., 4×10^{-9} cm.^{3}/sec. near 220 km. and 3×10^{-10} cm.^{3}/sec. near 300 km.* At the same time it must be realized that the coefficients during an aurora are not necessarily the same as those for normal conditions. Finally, η is known from quantal calculations by Bates & Massey (1943) to be about $1 \cdot 1 \times 10^{-15}$ cm.^{3}/sec. for attachment to atomic oxygen (assuming the resonance phenomena not to occur). The dissociative attachment process

---

* A discussion of the processes in an ionized atmosphere has been given by Bates & Massey (1946, 1947).

† For data on inelastic collision coefficients cf. Darrow (1932) and Mott & Massey (1933).

$$O_2 + e \rightarrow O^- + O \tag{A 2}$$

has a coefficient of about 10^{-10} cm.^{3}/sec. for electrons with energy of a few volts but does not occur for electrons in the passive group. Negative ions of nitrogen are not formed. Thus averaging over the constituents η can be taken as equal to 4×10^{-16} cm.^{3}/sec. Despite the uncertainties in the coefficients some interesting conclusions can be reached.

Using equation (A 1) table A 1, giving $n(eP)$ in units $\{\zeta/\alpha_e(1+\lambda)\}^{\frac{1}{2}}$ for a range of $n(eA)$ and $n(X)$, was compiled (a trifling task since the term containing η proved to be unimportant). In examining the results account must be taken of the unit in which the value of $n(eP)$ is quoted. $\{\zeta/\alpha_e\}^{\frac{1}{2}}$ can at once be seen to be about 3 if α_e is 1×10^{-8} cm.^{3}/sec. and about 20 if α_e is 3×10^{-10} cm.^{3}/sec. $\{1/(1+\lambda)\}^{\frac{1}{2}}$ is difficult to estimate reliably except at high altitudes. The attachment rate is $4 \times 10^{-16} n(X)$/electron so that in, say, $2 \cdot 5 \times 10^4$ sec. the number of negative ions formed is only $1 \times 10^{-11} n(X)$/electron. Hence in the region where $n(X)$ is 1×10^{11}/cm.^{3} or less the negative ion concentration will not normally have time to grow to be as large as the electron concentration and $\{1/(1+\lambda)\}^{\frac{1}{2}}$ can be assumed to be equal to unity. If $n(X)$ is much greater than 1×10^{11}/cm.^{3} the equilibrium negative ion concentration may, however, be established. This is controlled either by the associative detachment process

$$O^- + O \rightarrow O_2 + e, \tag{A 3}$$

or by the ionic recombination process

$$O^- + X^+ \rightarrow O' + X'. \tag{A 4}$$

Unfortunately the coefficients associated with neither of these is known at all reliably. But if that for the former K is taken to be about 10^{-16} to 10^{-17} cm.^{3}/sec. (averaged over the atmospheric constituents as usual) and that for the latter α_i to be about 10^{-10} cm.^{3}/sec.† (cf. Bates & Massey (1943)), then it can be seen that λ, whose equilibrium value is approximately given by the expression

$$\eta/(K + \alpha_i n(eP)), \tag{A 5}$$

cannot rise above 10 (and for the cases considered in the table is usually less) so that $\{1/(1+\lambda)\}^{\frac{1}{2}}$ is in general not smaller than $0 \cdot 3$. Hence the units $\{\zeta/\alpha_e(1+\lambda)\}^{\frac{1}{2}}$ vary from about 1 for high α_e and high λ to about 20 for low α_e and low λ.

It is clear from the results that almost all the auroral electrons are in the passive group. This is at variance with the statement sometimes made that their mean energy is of order 30 eV: but it is supported by the observational evidence that is available. Thus radio measurements have shown that the total electron concentration produced by aurorae must be at least comparable with that in the normal ionized layers. Even if an improbably low figure is taken for the excitation coefficient it is apparent that only a very small fraction of these can be in the active group as otherwise the intensity of the radiation that would of necessity ensue would be much too great.* Recent radar developments have led to the hope that a more detailed investigation of auroral conditions than hitherto possible can now be undertaken.

---

* Cf. 'The model ionosphere' table published by Bates & Massey (1946). These authors also give reasons for supposing that the recombination proceeds by an electronic, not ionic, reaction.

† It may well be greater: if this is the case λ becomes smaller.

TABLE A1

$n(X)$ (cm.^{3})	$n(eA)$ (cm.^{3})	$n(eP)$ units $\{\zeta/\alpha_e(1+\lambda)\}^{\frac{1}{2}}$ (cm.^{3})	$n(X)$ (cm.^{3})	$n(eA)$ (cm.^{3})	$n(eP)$ units $\{\zeta/\alpha_e(1+\lambda)\}^{\frac{1}{2}}$ (cm.^{3})
1×10^8	1	1×10^4	1×10^{12}	1×10^{-2}	1×10^5
	1×10^2	1×10^5		1	1×10^6
	1×10^4	1×10^6		1×10^2	1×10^7
1×10^{10}	1	1×10^5	1×10^{14}	1×10^{-2}	1×10^6
	1×10^2	1×10^6		1	1×10^7
	1×10^4	1×10^7		1×10^2	1×10^8

Valuable initial results have already been published by Lovell, Clegg & Ellyett (1947). Their interpretation has been discussed by Herlofson (1947) who deduced that in the particular aurora studied the electron concentration was greater than 4×10^4/cm.^{3} and considerably less than 3×10^7/cm.^{3}. This range is unfortunately rather wide. For the purpose of discussion, where a definite figure is desirable, 3×10^5/cm.^{3} seems a not unreasonable choice. Herlofson also deduced that the altitude was probably about 200 km. At this level there are slightly over 10^{11} atoms or molecules/cm.^{3}. Adapting for $n(eP)$ and $n(X)$ the values quoted and referring to the table it can be seen that rather less than 1/cm.^{3} is the value predicted for $n(eA)$. With an effective excitation coefficient of about 10^{-7} cm.^{3}/sec. or perhaps greater this concentration gives rise to an emission of some 10^4 photons/cm.^{3}/sec. which is just the rate usually assumed for aurorae. To provide a proper check on the theory precise electron concentration and intensity measurements are clearly needed. But even at present there is little doubt as to the correctness of the general conclusion that $n(eP)$ is very much greater than $n(eA)$. It is perhaps worth pointing out here that the treatment depends on the basic hypothesis that aurorae are produced by incoming particles. Some workers advocate the alternative hypothesis that aurorae are merely discharges caused by electric fields within the Earth's atmosphere. If this view were correct one would expect a higher proportion of energetic electrons than could be reconciled with observation.

As regards the energy distribution within the active group, little can be said at present as a number of unknown factors enter. It may, however, be remarked that the claim (already mentioned) that the electrons have an energy of about 30 eV is incorrect even if taken as applying to this group alone. It is true that if the electrons did have about this energy certain laboratory and auroral intensity measurements would be consistent with each other:* but consistency can be obtained more naturally in other ways—by assuming for example that $f(\epsilon)$, the distribution function, varies as ϵ^{-t} with t perhaps between 1 and 2 in the ϵ-range from 15 to 40 eV. In view of the earlier remarks on the possible unreliability of the measurements (§ 3·22) it is necessary to treat with reserve any $f(\epsilon)$ derived from them.

---

* It can readily be seen that the number of photons emitted/cm.^{3}/sec. would be many millions.

In connexion with the problem of understanding the emission of the negative bands of nitrogen from aurorae it is necessary to estimate the relative importance of the processes

$$N_2(X\,^1\Sigma_g^+) + e \rightarrow N_2^+(B\,^2\Sigma_u^+) + 2e, \qquad (A\,6)$$

and

$$N_2^+(X\,^2\Sigma_g^+) + e \rightarrow N_2^+(B\,^2\Sigma_u^+) + e. \qquad (A\,7)$$

The intensity resulting from the former will be denoted by $I(N_2)$, that from the latter by $I(N_2^+)$, and $I(N_2^+)/I(N_2)$ by r. If it is assumed that the nitrogen ions formed are mainly in the excited state and that they disappear mainly by recombination with electrons then it can be seen that

$$r \simeq \beta(N_2^+)\,n(eA)/\alpha_e(N_2^+)\,n(eP), \qquad (A\,8)$$

where $\beta(N_2^+)$ is the excitation coefficient associated with (A 7) and $\alpha_e(N_2^+)$ is the electronic recombination coefficient for nitrogen ions. The values of the coefficients are not known reliably, but their ratio $\beta(N_2^+)/\alpha_e(N_2^+)$ probably lies between 1 and 1000. These limits, despite their wideness, are sufficient for the present purpose. For, taken in conjunction with the data in the table, they show that r becomes comparable with unity only when $n(X)$ is very low indeed (less than perhaps 10^8/cm.^{3}).† Thus it would appear that process (A 7) is unimportant compared with process (A 6) except at extremely great altitudes.

REFERENCES

Barbier, D. 1947 *Ann. Géophys.* **3**, 227.
Barbier, D. 1948 *Report at Lyon Réunion.*
Bates, D. R. 1947 *Conference of Gassiot Committee* (Phys. Soc. London).
Bates, D. R. 1949 *Proc. Roy. Soc.* A (in the Press).
Bates, D. R. & Massey, H. S. W. 1943 *Phil. Trans.* A, **239**, 269.
Bates, D. R. & Massey, H. S. W. 1946 *Proc. Roy. Soc.* A, **187**, 261.
Bates, D. R. & Massey, H. S. W. 1947 *Proc. Roy. Soc.* A, **192**, 1.
Bates, D. R., Massey, H. S. W. & Pearse, R. W. B. 1947 *Conference of Gassiot Committee* (Phys. Soc. London).
Birge, R. J. 1926 *Bull. Nat. Res. Coun., Wash.*, **11**, 69.
Cabannes, J. & Dufay, J. 1946 *Ann. Géophys.* **2**, 290.
Darrow, K. K. 1932 *Electrical phenomena in gases.* Baltimore: Williams and Wilkens.
Dufay, J. 1943 *Ann. Astrophys.* **6**, 81.
Dufay, J. & Déjardin, G. 1946 *Ann. Géophys.* **2**, 249.
Dufay, J. & Dufay, M. 1948 *C.R. Acad. Sci., Paris*, **226**, 1208.

---

* Cf. Rypdal & Vegard (1940).

† If some of the nitrogen ions formed are initially in their ground state r is increased but a large factor cannot be introduced in this way: if some of the nitrogen ions disappear by processes other than electronic recombination (say by charge transfer collisions with oxygen atoms), r is decreased thereby strengthening the conclusion reached.

Dufay, J. & Tcheng Mao-Lin 1942 *Cah. Phys.* **8**, 51.
Dufay, M. & Dufay, J. 1947 *C.R. Acad. Sci., Paris*, **224**, 1834.
Gaydon, A. G. 1947 *Dissociation energies*. London: Chapman and Hall.
Ghosh, S. N. 1946 *Proc. Nat. Inst. Sci., India*, **12**, 405.
Götz, F. W. 1941 *Naturwissenschaften*, **29**, 690.
Herlofson, N. 1947 *Nature*, **160**, 867.
Herman, R., Herman, L. and Gauzit, J. 1942 *Cah. Phys.* **12**, 46.
Herzberg, G. 1928 *Z. Phys.* **49**, 761 and *Ann. Phys., Lpz.*, **86**, 191.
Herzberg, G. 1939 *Molecular spectra and molecular structure—Diatomic molecules*. New York: Prentice Hall.
Herzberg, G. & Herzberg, L. 1948 *Nature*, **161**, 283.
Hutchisson, E. 1930 *Phys. Rev.* **36**, 410.
Kallmann, H. & Rosen, B. 1929 *Z. Phys.* **58**, 52.
Langstroth, G. O. 1934 *Proc. Roy. Soc.* A, **146**, 166.
Langstroth, G. O. 1935 *Proc. Roy. Soc.* A, **150**, 371.
Lovell, A. C. B., Clegg, J. A. & Ellyett, C. D. 1947 *Nature*, **160**, 372.
Merton, T. R. & Pilley, J. G. 1925 *Phil. Mag.* **50**, 195.
Minnaert, M., Mulders, G. F. W. & Houtgast, J. 1940 *Photometric atlas of the solar spectrum*. Amsterdam: Schnabel.
Mitra, S. K. 1946 *Nature*, **157**, 692.
Morse, P. 1929 *Phys. Rev.* **34**, 57.
Mott, N. F. & Massey, H. S. W. 1933 *The theory of atomic collisions*. Oxford: Clarendon Press.
Ornstein, L. S. & van Wijk, W. R. 1928 *Z. Phys.* **49**, 315.
Pearse, R. W. B. & Gaydon, A. G. 1941 *The identification of molecular spectra*. London: Chapman and Hall.
Pettit, E. 1940 *Astrophys. J.* **91**, 159.
Poetker, A. H. 1927 *Phys. Rev.* **30**, 812.
Rayleigh, Lord 1922 *Proc. Roy. Soc.* A, **101**, 114 and 312.
Rayleigh, Lord 1930 *Proc. Roy. Soc.* A, **129**, 458.
Rypdal, B. & Vegard, L. 1940 *Geofys. Publ. Oslo*, **12**, no. 12.
Slipher, V. N. & Sommer, L. A. 1929 *Naturwissenschaften*, **17**, 802.
Smyth, H. D. & Arnott, E. G. F. 1930 *Phys. Rev.* **36**, 1021.
Spitzer, L. 1947 *Report at Yerkes Observatory Symposium on 'Planetary Atmospheres'*.
Sponer, H. 1935 *Molekül Spectren*. Berlin: Springer.
Stebbing, J., Whitford, A. E. & Swings, P. 1945 *Astrophys. J.* **101**, 39.
Størmer, C. 1939 *Terr. Mag., Atmos. Elect.* **44**, 7.
Swings, P. 1947 *Report at Yerks Observatory Symposium on 'Planetary Atmospheres'*.
Tawde, N. R. 1934 *Proc. Phys. Soc.* **46**, 324.
Vegard, L. 1932 *Geofys. Publ., Oslo*, **9**, no. 11.
Vegard, L. 1940 *Geofys. Publ., Oslo*, **12**, no. 14.
Vegard, L. & Kvifte, G. 1945 *Geofys. Publ., Oslo*, **16**, no. 7.
Vegard, L. & Tønsberg, E. 1935 *Geofys. Publ. Oslo*, **11**, no. 2.
Vegard, L. & Tønsberg, E. 1940 *Geofys. Publ., Oslo*, **13**, no. 1.
Wu, T. Y. 1944 *Phys. Rev.* **66**, 65.

The indefinite metric in relativistic quantum mechanics

By K. J. Le Couteur, *Manchester University*

(*Communicated by P. M. S. Blackett, F.R.S.—Received* 13 *July* 1948)

There are many problems of relativistic quantum mechanics which cannot be treated by working with a positive definite metric in the vector space of the fully quantized system. In this paper it is shown that, independently of the signs of energy and charge in the c number theory, any relativistic wave equation can be quantized and interpreted in terms of positive energies by use of an indefinite metric of suitable signature. A powerful method of obtaining positive definite transition probabilities from an indefinite metric is introduced and the concept of 'negative probability' is entirely avoided. The transition probabilities are defined as the squared matrix elements of a unitary matrix T, which is obtained from the solution of the Schrödinger equation by means of a special transformation. These techniques make it possible to treat relativistic wave equations which describe particles with several mass and spin states and to give a satisfactory account of the transition between two such mass states, a problem which is otherwise quite intractable.

The relationship of spin and statistics is discussed and it is concluded that, apart from the special cases of spin 0, $\frac{1}{2}$, 1, there is no necessary connexion between spin and statistics.

1. Introduction

In recent work on relativistic quantum mechanics (Dirac 1942, 1943; Pauli 1943) it has been found convenient to introduce a metric of indefinite sign in the vector space of the fully quantized system. When one applies such a formalism to problems in which states of both signs of the metric play an important part, the physical interpretation becomes rather awkward, and it has to be assumed that, in this domain, the only possible measurement is that of transition probabilities; this measurement requires only 'asymptotic' observation of the system. Following Heisenberg (1943), these transition probabilities will be described in terms of a 'scattering matrix' but the connexion with a Hamiltonian is not abandoned. Quite apart from this essential limitation on the physical interpretation, one meets the further difficulty that the quantities which, by analogy with ordinary quantum mechanics in a positive definite metric, one would like to interpret as transition probabilities are not all positive and are therefore not capable of immediate interpretation. The main contribution of this paper to the theory of quantum mechanics in an indefinite metric is to show that the transition probabilities can always be defined as the absolute squares of the matrix elements of a unitary matrix T, which is obtained from the solution of the Schrödinger equation by means of a special transformation defined in § 7. All requirements of the theory of probability are therefore satisfied and one needs to have no dealings whatever with 'negative probabilities'. The special transformation can be regarded as a generalized subtraction method but it is important to notice that it is not a perturbation method.

It is necessary to use an indefinite metric whenever one wishes to handle a relativistic wave equation involving spin greater than one. For it is well known that, if the metric is restricted to be positive definite, the quantization and interpretation of a relativistic particle theory is possible only when, in the c number theory, either the charge or the energy is positive definite. Recently Wild (1947) has shown that, apart from equations involving subsidiary conditions which cannot be quantized, the only wave equations meeting this requirement are those of spin 0, $\frac{1}{2}$, 1. Therefore the only possibility of progress beyond these equations lies in the development of a method of quantization and physical interpretation which does not require positive definite energy or charge. This problem is treated by working in an indefinite metric, the signature of which is defined in § 4, and by relating the physical interpretation to the T transformation. In the case of half-odd integral spins the problem of Fermi-Dirac quantization has already been solved (Bhabha 1947; Le Couteur 1948, referred to as A and B) but these papers do not settle the general question of the physical interpretation in an indefinite metric.

In §§ 2 and 3, which are introductory, a short account is given of the quantization and of the formulae which are used in § 4 to show how the signature of the metric in the Hilbert space is related to the signs of energy and charge in the c number theory. To establish the physical interpretation it is necessary to consider the interaction of the particles with another known field, for in the absence of interaction there are of course no problems to be answered. It is assumed that the particles are charged and so the interaction with the Maxwell field has been discussed, but the method is general and could equally well be applied to other interactions and to neutral particles. The Maxwell field is, however, of special interest because it involves peculiar supplementary conditions and it has to be verified that the present transformation procedure is consistent with these as well as with the equations of motion. As the method of expansion in series of eigenfunctions is itself modified by the interaction terms it has been thought desirable to carry through the whole theory in the presence of interaction. This procedure has the additional advantage of showing clearly that the T matrix subtraction formalism which is introduced is not necessarily a perturbation method.

The theory is presented in a form which applies equally to integral and half-odd integral spins and discussion of the relationship between spin and statistics is deferred till § 13. It is concluded that, apart from the cases of spin 0, $\frac{1}{2}$, 1 in which special simplifications exist, there is no necessary connexion between the spin and the statistics.

In §§ 11 and 12 the formalism is applied to wave equations describing particles which can appear in various mass and spin states and the problem of the transition between two different mass states is studied in detail.

2. C NUMBER THEORY

Use is made of relativistic notation with $(x_0, x_1, x_2, x_3) = (x_0, \mathbf{x}) = x$ as time and space co-ordinates and the metric tensor is taken in the form

$$g^{00} = -g^{11} = -g^{22} = -g^{33} = 1.$$

Suffixes μ, ν, σ run from 0 to 3, suffixes k, l from 1 to 3 and the three-dimensional scalar product is sometimes written as $\boldsymbol{\lambda}\mathbf{x} = \lambda_l x_l$.

Now consider a field of material particles which, in interaction with the electromagnetic field, is represented by the Lagrangian

$$L_P = \psi^\dagger(x) D(\alpha^\mu p_\mu - \chi)\psi(x). \tag{1}$$

Here $\overrightarrow{p_\mu} = i\overrightarrow{\partial_\mu} + eA_\mu$ and A_μ, satisfying $\partial^\mu A_\mu = 0$, is the four potential of the Maxwell field. χ is a constant. The four matrices α^μ define the structure and properties of the particle; since they are, in general, not Hermitian the matrix D, satisfying $D\alpha^\mu = (\alpha^\mu)^\dagger D$ and $D = D^\dagger$, must be introduced to make the Lagrangian real. D exists if, as we assume, the matrices α^μ are equivalent to their Hermitian conjugates $(\alpha^\mu)^\dagger$.

The invariance of L_P under the infinitesimal Lorentz transformation

$$x^\mu \rightarrow x^\mu + \epsilon^{\mu\nu}x_\nu, \quad \psi \rightarrow (1 + \tfrac{1}{2}\epsilon^{\mu\nu}I_{\mu\nu})\psi, \tag{2}$$

requires that the spin matrices $I_{\mu\nu}$ obey the commutation rules

$$(\alpha_\mu, I_{\nu\sigma}) = g_{\mu\nu}\alpha_\sigma - g_{\mu\sigma}\alpha_\nu. \tag{3}$$

In particular $(\alpha_0, I_{0l}) = \alpha_l$ which implies

$$(\alpha_0' \,|\, \alpha_l \,|\, \alpha_0') = 0. \tag{4}$$

It is convenient to choose the representation so that D commutes with α_0, I_{23}, I_{31}, I_{12} and anticommutes with α_l and I_{0l}; in the electron and meson theories and in Bhabha's generalization D is actually a function of α_0.

To L_P we add the Lagrangian L_R of the radiation field,

$$L_R = -\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu},$$

where

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{5}$$

By variation of the total Lagrangian (Weiss 1938; Pauli 1941; Bhabha 1945) we obtain the field equations

$$(\alpha^\mu p_\mu - \chi)\psi(x) = 0, \quad \psi^\dagger(x) D(\alpha^\mu p_\mu - \chi) = 0, \tag{6}$$

and

$$\frac{1}{4\pi}\partial_\mu F^{\mu\nu} = s^\nu, \tag{7}$$

where

$$s^\nu = e\psi^\dagger D\alpha^\nu\psi \tag{8}$$

is the charge current vector of the particles. The energy-momentum tensor is

$$T^{\mu\nu} = T^{\mu\nu}_P + T^{\mu\nu}_R,$$

where

$$T^{\mu\nu}_P = \psi^\dagger D\alpha^\mu i\partial^\nu\psi, \tag{9}$$

$$T^{\mu\nu}_R = -\frac{1}{4\pi}F^{\mu\sigma}\partial^\nu A_\sigma - g^{\mu\nu}L_R. \tag{10}$$

We must now discuss the procedure to be followed if the matrices α_μ are singular. Quite generally we suppose that they are of rank N and order $M \geqslant N$. Since D and α_0 commute and are Hermitian we may choose a representation in which they are both diagonal, the first N diagonal elements of α_0 being non-zero. In such a representation $(M-N)$ of the equations (6) do not involve time differentiations but are of type

$$_s(\alpha^l p_l - \chi)\,\psi(x) = 0, \quad \psi^\dagger(x)\,D(\alpha^l p_l - \chi)_s = 0 \qquad (\text{with } s > N). \tag{11}$$

The suffixes s, t which run from 1 to M are introduced to distinguish the different components of the wave function $\psi(x)$ and of the matrices α_μ. By use of (4), equations (11) may be written as

$$\chi\psi_s(x) = i\alpha^l_{st}\partial_l\psi_t(x) + e\alpha^l_{st}A_l\psi_t(x) \quad (\text{with } s > N \geqslant t).$$

We introduce a new wave function $\phi(x)$ defined by

$$\psi(x) \equiv \phi(x) + \theta(x), \quad \alpha_0\psi(x) = \alpha_0\phi(x), \tag{12}$$

with

$$\left.\begin{aligned} \phi_s(x) &= \psi_s(x) && (\text{for } s \leqslant N), \\ &= \chi^{-1}i\alpha^l_{st}\partial_l\psi_t(x) && (\text{for } s > N), \end{aligned}\right\} \tag{13a}$$

$$\left.\begin{aligned} \theta_s(x) &= 0, && (\text{for } s \leqslant N), \\ &= \chi^{-1}e\alpha^l_{st}A_l\psi_t(x) = \chi^{-1}e\alpha^l_{st}A_l\phi_t(x) && (\text{for } s > N). \end{aligned}\right\} \tag{13b}$$

At each instant of time these equations may be regarded as explicit definitions of the components $\phi_s(x)$, $\psi_s(x)$ for $s > N$ in terms of the N independent components with $s \leqslant N$. Suppose that, in the case $M > N$, the characteristic equation of α_0 is $\alpha_0 f(\alpha_0) = 0$, where $f(0) = 1$. This equation can be used to put ($13\,a, b$) into a form independent of the representation of α_0. We note the equations

$$f(\alpha_0)\,(\alpha^l p_l - \chi)\,\psi(x) = 0, \tag{14a}$$

$$f(\alpha_0)\,(i\alpha^l\partial_l - \chi)\,\phi(x) = 0, \tag{14b}$$

$$\psi^\dagger D(\alpha^l p_l - \chi + e\alpha^0 A_0)\,\theta(x) = 0, \tag{15a}$$

$$\theta^\dagger(x)\,D(\alpha^l p_l - \chi + e\alpha^0 A_0)\,\phi(x) = \chi^{-1}\phi^\dagger D e\alpha^m A_m f(\alpha_0)\,e\alpha^l A_l\phi. \tag{15b}$$

The momentum eigenstates of a free particle satisfy the equation

$$(i\alpha^\mu\partial_\mu - \chi)\,\phi(x) = 0$$

and are given by

$$\phi(x) = \phi(r, j, \mathbf{k})\,\mathrm{e}^{-ik_\mu x^\mu}, \tag{16}$$

where

$$k_\mu k^\mu = k_0^2 - \mathbf{k}^2 = (\chi/j)^2, \quad (\alpha^\mu k_\mu - \chi)\,\phi(r, j, \mathbf{k}) = 0, \tag{17}$$

and the ambiguity of sign is resolved by giving j and k_0 the same sign. By Lorentz transformation we may pass to the frame of reference in which the particle is at rest

$$k_\mu \rightarrow (\chi/j, 0, 0, 0), \quad \phi(r, j, \mathbf{k}) \rightarrow \phi(r, j, \mathbf{0}),$$

and

$$(\alpha^0\chi/j - \chi)\,\phi(r, j, \mathbf{0}) = 0, \quad \phi^\dagger(r, j, \mathbf{0})\,(D\alpha^0\chi/j - \chi) = 0.$$

The various orthogonal eigenvectors $\phi(r, j, \mathbf{0})$ corresponding to each non-vanishing eigenvalue j of α_0 are distinguished by the parameter r which may be used to specify the spin state of the particle in its proper frame of reference. N combinations of the parameters r and j are required. (4) shows that $\phi^\dagger(r, j, \mathbf{0})\,D\alpha^l\phi(r', j, \mathbf{0}) = 0$ and so, as in B, we may normalize the $\phi(r, j, \mathbf{k})$ to

$$(k_0)^{-1}\phi^\dagger(r, j, \mathbf{k})\,D\alpha_0\phi(r', j', \mathbf{k}) = D(r', j')\,\delta_{rr'}\delta_{jj'}, \tag{18}$$

where $D(r,j) = \pm 1$ is the eigenvalue of D associated with the eigenvector $\phi(r,j,\mathbf{0})$ of α_0 and D. (18) gives

$$\left\{\sum^{r,j} D\alpha_0 \phi(r,j,\mathbf{k}) D(r,j)(k_0)^{-1} \phi^\dagger(r,j,\mathbf{k}) D\alpha_0 - D\alpha_0\right\} \phi(r',j',\mathbf{k}) = 0.$$

There are N linearly independent vectors $\phi(r',j',\mathbf{k})$ and so we obtain

$$\sum^{r,j} D\alpha_0 \phi(r,j,\mathbf{k}) D(r,j)(k_0)^{-1} \phi^\dagger(r,j,\mathbf{k}) D\alpha_0 = D\alpha_0. \tag{19}$$

If $D\alpha_0$ is non-singular a factor $D\alpha_0$ may be removed from this equation.

3. QUANTIZATION

The remaining field equations involve time derivatives and may be written in the Hamiltonian form

$$i\frac{\partial \mathscr{E}}{\partial x_0} = (\mathscr{E}, H)_- \tag{20}$$

if we adopt the Hamiltonian $H = H_P + H_R$, where

$$H_P = -\int \Psi^*(x) D(\alpha^l p_l - \chi + e\alpha^0 A_0) \Psi(x)\, d^3x, \tag{21}$$

$$H_R = -\frac{1}{8\pi}\int\left\{\frac{\partial A_\mu}{\partial x_k}\frac{\partial A^\mu}{\partial x_k} + \frac{\partial A_\mu}{\partial x_0}\frac{\partial A^\mu}{\partial x_0}\right\}, \tag{22}$$

H_P and H_R are actually equal to the volume integrals of T^{00}_P and T^{00}_R. In (21) we have replaced the wave function $\psi(x)$ and its conjugate $\psi^\dagger(x)$ by adjoint operators $\Psi(x)$ and $\Psi^*(x)$ which obey the commutation rules, valid for $x_0 = x_0'$ only,

$$(\Psi_s(x), \Psi^*(x') D\alpha_0)_t)_\pm = ({}_t(D\alpha_0 \Psi(x')), \Psi^*_s(x))_\pm = \delta_{s,t}\delta(\mathbf{x}-\mathbf{x}').$$

In the case $M > N$ these rules are valid only in the representation in which α_0 is diagonal and for $s, t \leqslant N$. However, they may in all cases be written in a form independent of the representation of α_0 as

$$\left.\begin{aligned} (D\alpha_0 \Psi(x'), \Psi^*(x') D\alpha_0)_\pm &= D\alpha_0 \delta(\mathbf{x}-\mathbf{x}'), \\ (\Psi(x), \Psi(x'))_\pm = (\Psi^*(x), \Psi^*(x'))_\pm &= 0. \end{aligned}\right\} \tag{23}$$

The commutation rules for the potentials A_μ, which of course commute with the Ψ and Ψ^*, are

$$\begin{aligned} (\dot{A}_\mu(x), A_\nu(x'))_- &= 4\pi i g_{\mu\nu}\delta(\mathbf{x}-\mathbf{x}'), \\ (A_\mu(x), A_\nu(x'))_- &= (\dot{A}_\mu(x), \dot{A}_\nu(x'))_- = 0. \end{aligned} \tag{24}$$

Equations (23) and (24) are valid for $x_0 = x_0'$ only.

In this formalism we have to replace the equation $\partial^\mu A_\mu = 0$ by a supplementary condition

$$R\,|\rangle = 0, \tag{25}$$

which is a restriction, consistent with the equations of motion, on the vector representing the state of the fully quantized system. The equations of the material field do not give rise to any such conditions because equations (12) and (13), which hold also for the operators $\Psi(x)$, $\Phi(x)$, enable us to work entirely with the N independent components of $\Psi(x)$ or $\Phi(x)$.

Suppose that, at time x_0, we have a field $\Psi(x_0, \mathbf{x})$ which satisfies (14*a*). The associated field $\Phi(x)$ which must satisfy (14*b*) can be expressed, at time x_0, as a superposition of momentum eigenstates (16) and then (14*b*) is automatically satisfied. $\Psi(x_0, \mathbf{x})$ cannot be so expanded.

$$\left.\begin{aligned}\Phi(x_0,\mathbf{x}) &= (2\pi)^{-\frac{3}{2}}\sum_{r,j}\int\frac{d^3k}{k_0}D(r,j)\,U(x_0,r,j,\mathbf{k})\,\phi(r,j,\mathbf{k})\,\mathrm{e}^{-ik_lx^l},\\ \Phi^*(x_0,\mathbf{x}) &= (2\pi)^{-\frac{3}{2}}\sum\int\frac{d^3k}{k_0}D(r,j)\,U^*(x_0,r,j,\mathbf{k})\,\phi^\dagger(r,j,\mathbf{k})\,\mathrm{e}^{ik_lx^l}.\end{aligned}\right\}\tag{26}$$

$D(r,j) = \pm 1$ and has been defined in (18). The dynamical variables U, U^* obey the commutation rules

$$(U(x_0,r,j,\mathbf{k}),\,U^*(x_0,r',j',\mathbf{k}'))_\pm = k_0 D(r,j)\,\delta_{rr'}\delta_{jj'}\delta(\mathbf{k}-\mathbf{k}'),\tag{27}$$

from which we may derive (23) by use of (19) and (12). In these variables the Hamiltonian H_P becomes, by (12), (15*a*) and (15*b*),

$$\begin{aligned}H_P &= -\int\Psi^*(x)\,D(\alpha^lp_l-\chi+e\alpha^0A_0)\,\Phi(x)\,d^3x\\ &= -\int\Phi^*D(i\alpha^l\partial_l-\chi)\,\Phi(x)+H_{\text{int.}} = E_P+H_{\text{int.}},\end{aligned}$$

$$E_P = \sum_{r,j}\int\frac{d^3k}{k_0}D(r,j)\,U^*(rj\mathbf{k})\,U(rj\mathbf{k})\,k_0,\tag{28a}$$

$$H_{\text{int.}} = -\int\Phi^*De\alpha^\mu A_\mu\Phi d^3x-(\chi)^{-1}\int\Phi^*De\alpha^mA_mf(\alpha_0)\,e\alpha^lA_l\Phi\,d^3x.\tag{28b}$$

The total charge is

$$\begin{aligned}Q &= e\int\Psi^*D\alpha^0\Psi\,d^3x = e\int\Phi^*D\alpha^0\Phi\,d^3x\\ &= e\sum_{r,j}\int\frac{d^3k}{k_0}D(r,j)\,U^*(r,j,\mathbf{k})\,U(r,j,\mathbf{k}).\end{aligned}\tag{29}$$

The total momentum of the particles is, by (9),

$$\begin{aligned}G^l_P &= \int\Phi^*D\alpha^0i\partial^l\Phi\,d^3x\\ &= \sum\int\frac{d^3k}{k_0}D(r,j)\,U^*(r,j,\mathbf{k})\,U(r,j,\mathbf{k})\,k^l.\end{aligned}\tag{30}$$

We treat the radiation field similarly. At each instant of time A_μ and $\dot{A}_\mu$ are regarded as independent dynamical variables (Dirac 1947), and may be expressed as Fourier expansions

$$\left.\begin{aligned}A_\mu(x_0,\mathbf{x}) &= \frac{1}{2\pi}\int\{A^*_\mu(x_0,\boldsymbol{\lambda})\,\mathrm{e}^{i\lambda_lx^l}+A_\mu(x_0,\boldsymbol{\lambda})\,\mathrm{e}^{-i\lambda_lx^l}\}\frac{d^3\lambda}{|\boldsymbol{\lambda}|},\\ \dot{A}_\mu(x_0,\mathbf{x}) &= \frac{i}{2\pi}\int\{A^*_\mu(x_0,\boldsymbol{\lambda})\,\mathrm{e}^{i\lambda_lx^l}-A_\mu(x_0,\boldsymbol{\lambda})\,\mathrm{e}^{-i\lambda_lx^l}\}\,d^3\lambda.\end{aligned}\right\}\tag{31}$$

The commutation rule (24) becomes

$$(A_\mu(x_0, \boldsymbol{\lambda}), A_\nu^*(x_0, \boldsymbol{\lambda}))_- = -g_{\mu\nu} \,|\, \boldsymbol{\lambda} \,|\, \delta(\boldsymbol{\lambda} - \boldsymbol{\lambda}'). \tag{32}$$

In these variables we obtain from (22)

$$H_R = -\frac{1}{2}\int d^3\lambda \{A_\mu^*(\boldsymbol{\lambda})\, A^\mu(\boldsymbol{\lambda}) + A_\mu(\boldsymbol{\lambda})\, A^{*\mu}(\boldsymbol{\lambda})\}. \tag{33}$$

The momentum of the radiation field may be derived from (10) as

$$\begin{aligned} G_R^l &= -\frac{1}{4\pi}\int \frac{\partial A_\mu}{\partial x_0}\frac{\partial A_\mu}{\partial x_l} d^3x \\ &= -\frac{1}{2}\int \frac{d^3\lambda}{|\,\boldsymbol{\lambda}\,|} \lambda^l \{A_\mu^*(\boldsymbol{\lambda})\, A^\mu(\boldsymbol{\lambda}) + A_\mu(\boldsymbol{\lambda})\, A^{*\mu}(\boldsymbol{\lambda})\}. \end{aligned} \tag{34}$$

At this stage it is convenient to enclose the whole system in a box of volume Ω. We have now a discrete set of momentum eigenstates $a, b, c, \ldots$ of the Ψ field, these are specified as before by the parameters $r(a), j(a), \mathbf{k}(a)$. At the same time we make the trivial contact transformation from the dynamical variables U, A_μ to dynamical variables V, B_μ, which in the absence of interaction between the two fields are constants of the motion.

$$\Phi(x) = \Omega^{-\frac{1}{2}} \sum_a V(x_0, a)\, s_a \,|\, k_0(a) \,|^{-\frac{1}{2}} \phi(a)\, \mathrm{e}^{-ik_\mu x^\mu}, \tag{35}$$

$$V(x_0, a)\, \mathrm{e}^{-ik_0 x^0} = U(x_0, a)\, (2\pi)^{\frac{3}{2}}\, (\Omega \,|\, k_0 \,|)^{-\frac{1}{2}}, \tag{36}$$

where $k_0(a)$ is defined by (17). We introduce the notation $\operatorname{sgn} j = j/|\,j\,|$ and define

$$s(a) = D(r(a), j(a)) \operatorname{sgn} j(a). \tag{37}$$

The commutation rule (27) is now

$$(V(x_0, a), V^*(x_0, b))_\pm = s(a)\, \delta_{a,b}. \tag{38}$$

Similarly with $\lambda_0 = |\,\boldsymbol{\lambda}\,|$ we define $B_\mu(x_0, \boldsymbol{\lambda})$ by

$$B_\mu(x_0, \boldsymbol{\lambda})\, \mathrm{e}^{-i\lambda_0 x^0} = A_\mu(x_0, \boldsymbol{\lambda})\, (2\pi)^{\frac{3}{2}}\, (\Omega\lambda_0)^{-\frac{1}{2}} \tag{39}$$

and then

$$(B_\mu(x_0, \boldsymbol{\lambda}), B_\nu^*(x_0, \boldsymbol{\lambda}'))_- = -g_{\mu\nu}\, \delta_{\boldsymbol{\lambda}, \boldsymbol{\lambda}'}. \tag{40}$$

The Hamiltonian of the combined system is now just $H_{\text{int.}}$ defined by (28). The supplementary condition (25) is equivalent to

$$\lambda^\mu B_\mu(\boldsymbol{\lambda}) \,|\,) = 0, \quad \lambda^\mu B_\mu^*(\boldsymbol{\lambda}) \,|\,) = 0. \tag{41}$$

We define the quantum numbers of the Maxwell field as

$$n_0(\boldsymbol{\lambda}) = B_0(\boldsymbol{\lambda})\, B_0^*(\boldsymbol{\lambda}), \quad n_l(\boldsymbol{\lambda}) = B_l^*(\boldsymbol{\lambda})\, B_l(\boldsymbol{\lambda}). \tag{42}$$

Then, if the suffix ϵ denotes a transverse polarization, the expressions (33) and (34) for the total energy and momentum of the Maxwell field become

$$(E_R, G_R^l) = \Sigma n_\epsilon\, (\boldsymbol{\lambda})\, (\lambda^0, \lambda^l). \tag{43}$$

The summation is confined to the transverse polarizations because, for any state satisfying (41), the contributions of the two longitudinal polarizations cancel. As usual we choose the representation so that the eigenvalues of each $n_\mu(\boldsymbol{\lambda})$ consist of the set of all positive integers, including 0.

4. THE SIGNATURE OF THE METRIC

It is now necessary to distinguish between Bose and Fermi-Dirac statistics.

(*a*) *Bose statistics*

According to (28) and (38) we have

$$E_P = \Sigma D_a V_a^* V_a \,|\, k^0 \,|, \quad D_a V_a V_a^* - D_a V_a^* V_a = \operatorname{sgn} j_a. \tag{44}$$

We define the quantum numbers of the material field as

$$\left.\begin{aligned} D_a V_a^* V_a &= n_a, \quad D_a V_a V_a^* = n_a + 1 \qquad (\text{for } \operatorname{sgn} j_a = +1), \\ D_a V_a V_a^* &= n_a, \quad D_a V_a^* V_a = n_a + 1 \qquad (\text{for } \operatorname{sgn} j_a = -1). \end{aligned}\right\} \tag{45}$$

As eigenvalues of each n_a we wish to have the set of all positive integers and zero. In the case $D_a = -1$ we must (Pauli 1943) quantize the V_a, V_a^* with an indefinite metric and we proceed to determine the metric operator η in terms of the dynamical variables n_a and the parameters D_a. We may take as basic states the simultaneous eigenstates $(n'_a, n'_b \ldots ; n'_\mu(\boldsymbol{\lambda}) \ldots |$ of the commuting observables n_a, $n_b \ldots n_\mu(\boldsymbol{\lambda}) \ldots$. In this representation the matrix elements of V_a, V_a^* are defined by

$$\left.\begin{aligned} (n'_a, n'_b \ldots ; n'_\mu(\boldsymbol{\lambda}) \ldots | V_a |) &= (n'_a | V_a | n'_a + 1)\,(n'_a + 1, n'_b \ldots ; n'_\mu(\boldsymbol{\lambda}) \ldots |) \\ (n'_a + 1, n'_b \ldots ; \ldots | V_a^* |) &= (n'_a + 1 | V_a^* | n'_a)\,(n'_a, n'_b, \ldots ; \ldots |) \end{aligned}\right\} \quad (\text{for } j > 0), \tag{46}$$

and for $j < 0$ (45) shows that in (46) we must interchange V_a and V_a^*. The scalar product of two states $(p|$ and $|q)$ is defined as

$$(p | \eta | q) = \sum_{n'_a \ldots n'_\mu(\lambda)} (p | n'_a, \ldots ; n'_\mu \ldots)\, \eta(n'_a, n'_b \ldots)\, (n'_a, n'_b \ldots ; n'_\mu \ldots | q), \tag{47}$$

η does not depend on the quantum numbers of the Maxwell field and for $B_\mu(\boldsymbol{\lambda})$, $B_\mu^*(\boldsymbol{\lambda})$ we have the well-known matrix elements

$$(n'_\mu(\lambda) | B_\mu(\boldsymbol{\lambda}) | n'_\mu(\boldsymbol{\lambda}) + 1) = (n'_\mu(\boldsymbol{\lambda}) + 1 | B_\mu^* | n'_\mu(\boldsymbol{\lambda})) = \sqrt{(n'_\mu(\boldsymbol{\lambda}) + 1)}. \tag{48}$$

In Pauli's formalism, the adjoint P^* of an operator P is related to the Hermitian conjugate $P^\dagger$ by

$$P^* = \eta^{-1} P^\dagger \eta. \tag{49}$$

By substitution of the matrix elements (46) into (45) we obtain, after use of (49),

$$\left.\begin{aligned} (n'_a | V_a | n'_a + 1) &= \sqrt{(n'_a + 1)}, \quad (n'_a + 1 | V_a^* | n'_a) = D_a \sqrt{(n'_a + 1)} \quad (\text{for } \operatorname{sgn} j_a = 1), \\ (n'_a | V_a^* | n'_a + 1) &= \sqrt{(n'_a + 1)}, \quad (n'_a + 1 | V_a | n'_a) = D_a \sqrt{(n'_a + 1)} \quad (\text{for } \operatorname{sgn} j_a = -1), \end{aligned}\right\} \tag{50}$$

and

$$\left.\begin{aligned} \eta(n'_a, n'_b \ldots) &= \eta_a(n'_a)\, \eta_b(n'_b) \ldots, \\ \eta_a(n'_a) &= D_a \eta_a(n'_a - 1) = (D_a)\, n'_a. \end{aligned}\right\} \tag{51}$$

with

Omitting the zero point contribution of one quantum for each eigenstate with $\operatorname{sgn} j = -1$, we obtain for the total energy, momentum and charge

$$E = \Sigma n_a \,|\, k^0(a) \,| + H_{\text{int.}} + \Sigma n_\epsilon(\boldsymbol{\lambda})\, \lambda^0 = E_0 + H_{\text{int.}}, \tag{52a}$$

$$G^l = \Sigma n_a(k^l(a) \operatorname{sgn} j(a)) + \Sigma n_\epsilon(\boldsymbol{\lambda})\, \lambda^l, \tag{52b}$$

$$Q = e\Sigma n_a \operatorname{sgn} j(a). \tag{52c}$$

(*b*) *Fermi-Dirac statistics*

The quantization has been given in A and B but the main results are repeated. We introduce auxiliary operators v_a, v_a^* which obey the commutation rules

$$\left.\begin{array}{c} (v_a, v_a^*)_+ = s_a, \quad v_a v_a = v_a^* v_a^* = 0, \\ v_a, v_a^* \text{ commute with } v_b, v_b^* \text{ (for } a \neq b). \end{array}\right\} \tag{53}$$

Then
$$n_a = s_a v_a^* v_a$$

has eigenvalues 0 and 1 only. We adopt a definite but arbitrary order for the $a, b, \ldots, r, \ldots$ and, in the representation with $n_a, n_b, \ldots, n_r, \ldots$ diagonal, the representation of v, v^* is

$$\left.\begin{array}{l} (n'_a \ldots n'_r \ldots; n'_\mu \ldots \,|\, v_r \,|) = \begin{cases} s_r(n'_a \ldots (n'_r + 1) \ldots; n'_\mu \ldots |) & (\text{if } n'_r = 0) \\ 0 & (\text{if } n'_r = 1) \end{cases}, \\ (n'_a \ldots n'_r \ldots; n'_\mu \ldots \,|\, v_r^* \,|) = \begin{cases} 0 & (\text{if } n'_r = 0) \\ (n'_a \ldots (n'_r - 1) \ldots; n'_\mu \ldots \,|) & (\text{if } n'_r = 1) \end{cases}. \end{array}\right\} \tag{54}$$

The scalar product of two states $(p\,|$ and $|\,q)$ is still defined by (47) with

$$\eta(n'_a, n'_b \ldots) = \eta_a(n'_a)\, \eta_b(n'_l) \ldots \quad \text{and} \quad \eta_r(1) = s_r, \quad \eta_r(0) = 1. \tag{55}$$

Since $(1 - 2n_r)$ anticommutes with v_r, v_r^* and commutes with all other v_s, v_r^* we may define the V_r, V_r^* in terms of the v_r, v_r^* as

$$\left.\begin{array}{r} V_r = (1-2n_a)(1-2n_b) \ldots (1-2n_{r-1})\, v_r, \\ V_r^* = V_r^*(1-2n_{r-1}) \ldots (1-2n_a), \end{array}\right\} \tag{56}$$

and
$$n_r = s_r v_r^* v_r = s_r V_r^* V_r. \tag{57}$$

The total energy, momentum and charge given by (28) to (30) become

$$E = \Sigma n_a k^0(a) + H_{\text{int.}} + \Sigma n_\epsilon \lambda^0 = E^0 + H_{\text{int.}}, \tag{58a}$$

$$G^l = \Sigma n_a k^l(a) + \Sigma n_\epsilon \lambda^l, \tag{58b}$$

$$Q = e\Sigma n_a. \tag{58c}$$

5. THE SCHRÖDINGER EQUATION

We now pass to the Schrödinger picture in which, dropping the suffix 'int.', we have a wave equation

$$i\frac{d}{dx_0}\,|\, x_0) = H \,|\, x_0), \tag{59}$$

with the formal solution
$$|\, t) = S \,|\, 0), \tag{60}$$

where the linear operator $S = S(t)$ is a function of the Hamiltonian given by

$$S = \exp\left(-i\int_0^t H\,dx_0\right).$$

In the case of a positive definite metric x_0 represents the observer's time co-ordinate and the squared amplitudes of the components of $|\,x_0)$ have a formal interpretation as probabilities, at time x_0, for the various possible states of the physical system. It is not strictly necessary to assume so many points of contact between the mathematical formalism and the physical world; indeed the discussion by Heisenberg (1943) and Møller (1947) shows that actual observations of physical systems involve only: (i) the asymptotic behaviour at ∞ in the observer's space-time system of the wave functions in collision, emission and absorption processes; this is conveniently described by means of a unitary 'scattering matrix', (ii) the discrete energy levels of closed stationary states of the system. It is important to notice that the observable quantities (i) and (ii) do not depend in any essential way on the time co-ordinate x_0.

In the case of an indefinite metric the number of points of contact between the mathematical formalism and the physical world is necessarily more limited but we shall see that it remains possible to calculate the quantities (i) and (ii) required by Heisenberg and Møller. However, there is no immediate interpretation of the squared amplitudes of the components of $|\,x_0)$ and even x_0 itself can no longer consistently be interpreted as the observer's time co-ordinate but must be regarded as a related mathematical variable. At the end of § 8 we shall make a suggestion as to the significance of x_0.

The whole theory has been developed without approximation by variation of a Lagrangian and we have therefore rigorous conservation of E, G, Q. However, for the further development of the theory conservation of E_0 is required. Now in the calculation of the transition processes (i) it is the limiting form, supposed existent, of (60) when 0 and t correspond to $x_0 = -\infty$ and $x_0 = +\infty$ that is involved. In such problems either $|\,0)$ or $|\,t)$ or both must belong to the continuous spectrum and then we assume that the expansion of $|\,0)$ and $|\,t)$ in terms of free particle eigenstates can be carried through in such a way that we have conservation of E_0 as well as conservation of E.

We have also the subsidiary conditions (25) or (41) which are known to be consistent with the equations of motion and therefore must commute with S.

6. The S-transformation

In place of the notation $(n'_a, n'_b \ldots;\ n'_\mu(\lambda) \ldots\,|$ of § 4 it will, in the following, be sufficient to label the basic states as $i, j \ldots$, etc. Similarly we write p_i for the i component $(i\,|\,p)$ of a vector $|\,p)$. In this ordinary matrix notation the invariant 'squared length' of $|\,p)$ is, according to (47),

$$(p\,|\,\eta\,|\,p) = \sum^{i} \bar{p}_i \eta_i p_i = \sum^{i} \pm |\,p_i\,|^2. \tag{61}$$

It is not possible to specify in an invariant way the order in which the terms of (61) are to be taken. We therefore admit only vectors p for which the infinite series (61) is absolutely convergent and therefore independent of the ordering. The absolute convergence of the scalar product

$$(p \mid \eta \mid q) = \sum^{i} \pm \bar{p}_i q_i, \tag{62}$$

is a further consequence of this assumption.

The transformation (60), $|\,t) = S\,|\,0)$ conserves the normalization (61), provided that

$$\eta^{-1} S^{\dagger} \eta S = 1 \quad \text{or} \quad S^{*} S = 1. \tag{63}$$

This implies $\sum^{i} \eta_i \eta_j \,|\, S_{ij} \,|^2 = 1$ so that, if we suppose that $|\,0)$ and $|\,t)$ represent the initial and final physical conditions of the system, then the formal analogy with the case of a positive definite metric suggests that we should interpret $\eta_i \eta_j \,|\, S_{ij} \,|^2$ as the probability of a transition from state i to state j. This 'transition probability' is negative if η_i and η_j are of different sign, however the sum is correctly normalized to 1. A possible method of procedure (Dirac 1943) is to assume that these 'transition probabilities' are not individually observable and to combine them in such a way as to obtain positive quantities. In Dirac's papers this assumption is required by the existence of redundant variables; it was made in B because it is physically reasonable to suppose that at extreme relativistic energies the various mass states are not experimentally distinguishable.

This procedure is inadequate if all the transition probabilities are regarded as individually observable and it becomes necessary to find a method of connecting $|\,0)$ and $|\,t)$ through a unitary matrix T. The absolute squares of the matrix elements of T may then be interpreted as physically observable transition probabilities.

7. THE T-TRANSFORMATION

Take a block representation which separates the positive and negative eigenvalues of η

$$\eta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad |\,) = \begin{pmatrix} x \\ y \end{pmatrix}. \tag{64}$$

The 'squared length' (61) becomes

$$(|\,\eta\,|) = x^{\dagger}x - y^{\dagger}y, \tag{65}$$

and the assumption of absolute convergence implies that $x^{\dagger}x$ and $y^{\dagger}y$ are separately convergent.

We write the transformation (60) as

$$|\,0) \rightarrow |\,t) = \begin{pmatrix} v \\ w \end{pmatrix} = S\,|\,0) = S\begin{pmatrix} x \\ y \end{pmatrix}, \tag{66}$$

and, since S conserves the normalization

$$v^{\dagger}v - w^{\dagger}w = x^{\dagger}x - y^{\dagger}y. \tag{67}$$

Hence

$$v^{\dagger}v + y^{\dagger}y = x^{\dagger}x + w^{\dagger}w. \tag{68}$$

Therefore, associated with the transformation S, we have a unitary transformation T

$$|p) = \begin{pmatrix} x \\ w \end{pmatrix} \rightarrow q = \begin{pmatrix} v \\ y \end{pmatrix} = T\begin{pmatrix} x \\ w \end{pmatrix}, \tag{69}$$

which leaves invariant the positive definite 'squared length' (68). Now from (64) and (66), it follows that

$$\begin{pmatrix} x \\ w \end{pmatrix} = \{\tfrac{1}{2}(1+\eta) + \tfrac{1}{2}(1-\eta)\, S\}\begin{pmatrix} x \\ y \end{pmatrix},$$

$$\begin{pmatrix} v \\ y \end{pmatrix} = \{\tfrac{1}{2}(1+\eta)\, S + \tfrac{1}{2}(1-\eta)\}\begin{pmatrix} x \\ y \end{pmatrix},$$

therefore, by (69),

$$T = \{1+S-\eta(1-S)\}\{1+S+\eta(1-S)\}^{-1}. \tag{70}$$

Provided that $\det(1+S) \neq 0$, we have

$$T = \left\{1-\eta.\frac{(1-S)}{1+S}\right\}\left\{1+\eta.\frac{1-S}{1+S}\right\}^{-1} = \frac{1-\eta K}{1+\eta K}, \tag{71}$$

where

$$K = \frac{1-S}{1+S} \quad \text{and} \quad S = \frac{1-K}{1+K}. \tag{72}$$

(63) is satisfied if

$$K^{\dagger}\eta + \eta K = 0 \quad \text{or.} \quad (\eta K)^{\dagger} + (\eta K) = 0. \tag{73}$$

Independently of the particular representation (64) of η, equations (70) to (73) can be used to derive a unitary transformation T from an S transformation.

The expectation value $(|\, \eta H_0\, |)$ of an observable H_0 is conserved if $S^{\dagger}\eta H_0 S = \eta H_0$ or, by (63), $S^{-1}H_0 S = H_0$. This condition is equivalent to

$$(H_0, S)_{-} = 0 \quad \text{or} \quad (H_0, K)_{-} = 0, \tag{74}$$

(74) implies that

$$(H_0, \eta K)_{-} = (H_0, \eta)_{-} K.$$

Therefore if the S transformation conserves the expectation value of an observable H_0 which commutes with η, then the associated T transformation also conserves H_0. Further the reality conditions $(\eta H_0)^{\dagger} = \eta H_0$ and $H_0^{\dagger} = H_0$ are equivalent. These results are particularly simple in the representation (64). If

$$H_0 = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$

then from the conservation law

$$v^{\dagger}Av - w^{\dagger}Bw = x^{\dagger}Ax - y^{\dagger}By, \tag{75}$$

we infer

$$v^{\dagger}Av + y^{\dagger}By = x^{\dagger}Ax + w^{\dagger}Bw. \tag{76}$$

It is obvious that in problems in which states of both signs of the metric play an important part the T transformation can only be used to connect an initial and a final state and not to provide a continuous description of the evolution of the system. There is no Schrödinger equation in the T-picture, we are limited to the calculation of the asymptotic values of wave functions. However, this is just the position that arises, quite apart from any considerations of indefinite metrics, in the theory of 'observable quantities' developed by Heisenberg and Møller (1947). Our theory of course differs from that of these authors in that we have a formal method of calculating T from the Hamiltonian.

8. PHYSICAL INTERPRETATION

We saw that the S transformation conserves the operators Q, G, E_0. It is obvious from the definition in § 4 of these operators that they commute with η and are therefore conserved in the T transformation. Similarly the subsidiary conditions (25) or (41) which commute with S and with η must also commute with T. In a theory with a positive definite metric the physical interpretation may be based on the assumption that the S transformation represents an observable transfer of energy momentum and charge from one state of the unperturbed system to another such state. In the case of the indefinite metric, the T transformation is introduced to represent the same transfer, subject to equivalent conservation laws, but in a form suitable for physical interpretation in terms of positive definite transition probabilities $|T_{ij}|^2$. The relationship of (76) to (75) and of (68) to (67) shows that we have here a subtraction method rather analogous to the 'hole' theory of positrons.

In the first-order approximation to S and T one takes

$$S = 1 - i\int_0^t H\,dx_0 = 1 - 2K \quad \text{and} \quad T = 1 - 2\eta K, \tag{77}$$

and in this approximation the transition probability $|T_{ij}|^2$ may be replaced by $|S_{ij}|^2$. It is not so in the higher approximations.

The dependence of the matrix elements of V, V^*, B, B^* and so of H on the quantum numbers n is defined by (48) and (50) and leads, in the approximation (77), to Einstein's laws of radiation, which in this use of the indefinite metric need not be introduced as additional assumptions.

We have defined our basic states $|n'_a\ldots)$ in a Lorentz invariant way and so, when $|0)$ and $|t)$ refer in the S picture to $x_0 = -\infty$ and $x_0 = +\infty$, all the matrix elements of S behave under Lorentz transformation as scalars. The definition of η is invariant and therefore the matrix elements of T are also scalars. The description of transition probabilities in terms of the T matrix is therefore a relativistic one.

We have now shown that our T satisfies all the mathematical requirements for a scattering matrix and we may therefore establish the physical interpretation of our formalism by asserting that *T is the observable scattering matrix of the quantized system*. (69) then shows that the squared amplitudes of the components of the wave functions $\begin{pmatrix} x \\ w \end{pmatrix}$ and $\begin{pmatrix} v \\ y \end{pmatrix}$ may be interpreted as observable probabilities for the various states of the physical system in the observers distant past and distant future.

In going from the S to the T picture we have had to mix up the wave functions $\begin{pmatrix} x \\ y \end{pmatrix} = |0)$ and $\begin{pmatrix} v \\ w \end{pmatrix} = |t)$. It is necessary to assume that these do not directly describe states in the observers distant past and distant future but rather refer to the starting and finishing points of the formal mathematical calculation of the matrix S. We see that x_0 can no longer be interpreted as the observer's time co-ordinate. It is tempting to speculate that our $|\,x_0)$ is the wave function that would arise by passing from a 'many-time' picture (Chang 1947) to a Schrödinger picture with a single time co-ordinate by putting $t_r = t$ for some particles and $-t_r = t$ for others. However, as there is no satisfactory formulation of the 'many-time' picture for interacting particles this must remain a conjecture.[†]

The necessity for distinguishing between 'initial' and 'final' data and the observer's past and future does not arise in the steady state treatment of scattering problems. We shall show that, in such problems, one obtains a unitary scattering matrix by modifying the definition of ingoing and outgoing waves.

9. THEORY OF SCATTERING

We now show that a suitable formulation of the well-known (Pauli 1947; Wentzel 1947) stationary state theory of scattering leads to S and T matrices related to each other by equations like (71) and (72).

We combine a complete set of solutions of the Schrödinger equation

$$(H_0 + H - E_0)\,|\,) = 0, \tag{78}$$

to form a matrix Φ such that the result of multiplying Φ into any eigenvector of the Hamiltonian H_0 of the system without interaction, is an eigenvector of the total Hamiltonian $H_0 + H$ with the same eigenvalue E_0. Φ is not fully determined as it can always be replaced by $\Phi\Lambda$, where Λ is any matrix which commutes with H_0. In the momentum representation Φ satisfies the equation

$$(E_0 - E_k)\Phi_{k0} = (H\Phi)_{k0}. \tag{79}$$

We solve (79) by taking $\Phi_{k0} = \delta_{k0} + F_{k0} = \delta_{k0} - \dfrac{1}{E_k - E_0} f_{k0}$, (80)

and then (79) requires $f = H + HF.$ (81)

With the definition (49) of the adjoint we have, since $H = H^*$,

$$f^* = H + F^*H. \tag{82}$$

Then following Pauli's treatment we obtain

$$f^* - f + f^*F - F^*f = 0. \tag{83}$$

[†] With Bose statistics it is possible, and at first seems simpler, to quantize the theory in a positive definite metric. We have then two kinds of particles, a and b say, of positive and negative energy respectively and with positive quantum numbers $m_a = V_a^* V_a$ and $m_b = V_b^* V_b$. The matrix element $(m_a', m_b' \ldots \,|\, S \,|\, m_a'', m_b'' \ldots)$ of the unitary scattering matrix S has then to be interpreted as representing an observable transition between physical states with particles $(m_a', m_b'', \ldots)$ and $(m_a'', m_b', \ldots)$ of positive energy. The scattering matrix T would be defined by

$$(m_a', m_b'' \ldots \,|\, T \,|\, m_a'', m_b' \ldots) = (m_a', m_b' \ldots \,|\, S \,|\, m_a'', m_b'' \ldots).$$

This procedure fails except in first approximation because this T is not unitary so that the observed cross-sections would not obey the ordinary laws of probability.

In Pauli's notation $\mathbf{f}$ is a submatrix of f belonging to a single energy shell. Then from (83) we obtain

$$\mathbf{f}^* = \mathbf{f} \quad \text{and} \quad \mathbf{f}^\dagger \boldsymbol{\eta} = \boldsymbol{\eta}\mathbf{f}. \tag{84}$$

If we write

$$\delta_{k0} = \delta(E_k - E_0)\, \mathbf{1}_{K0}, \tag{85}$$

and use the notation $\delta_\pm(E_k - E_0) = \tfrac{1}{2}\delta(E_k - E_0) \mp \dfrac{1}{2\pi i(E_k - E_0)}$, (86)

we may express (80) as

$$\Phi_{k0} = \delta_+(E_k - E_0)\{\mathbf{1}_{K0} + i\pi f_{k0}\} + \delta_-(E_k - E_0)\{\mathbf{1}_{K0} - i\pi f_{k0}\}. \tag{87}$$

We are interested in the coefficients $\mathbf{1} \pm i\pi\mathbf{f}$ of δ_+ and δ_- when the state k has the same energy as the state 0; because of the indeterminacy of Φ only their ratio is significant. The coefficient of the 'outgoing wave' is obtained from that of the 'incoming wave' by multiplication with a scattering matrix

$$\mathbf{S} = \frac{\mathbf{1} + i\pi\mathbf{f}}{\mathbf{1} - i\pi\mathbf{f}}, \tag{88}$$

and (84) shows that

$$\mathbf{S}^*\mathbf{S} = \mathbf{S}\mathbf{S}^* = \mathbf{1}. \tag{89}$$

Now we are working in the momentum representation in which η is a diagonal matrix. Therefore we may introduce the operators

$$\delta^\eta_\pm(k, 0) = \tfrac{1}{2}\delta(E_k - E_0) \mp \frac{1}{2\pi i(E_k - E_0)}\, \eta_k, \tag{90}$$

and then we break up Φ as

$$\begin{aligned}\Phi_{k0} &= \delta(E_k - E_0)\, \mathbf{1}_{K0} - \frac{1}{E_k - E_0}\, \eta_k(\eta f)_{k0} \\ &= \delta^\eta_+(k, 0)\{\mathbf{1}_{K0} + i\pi(\eta f)_{k0}\} + \delta^\eta_-(k, 0)\{\mathbf{1}_{K0} - i\pi(\eta f)_{k0}\}.\end{aligned} \tag{91}$$

When $E_k = E_0$ the ratio of the coefficients $\mathbf{1} \pm i\pi\boldsymbol{\eta}\mathbf{f}$ of δ^η_+ and δ^η_- is given by

$$\mathbf{T} = \frac{\mathbf{1} + i\pi\boldsymbol{\eta}\mathbf{f}}{\mathbf{1} - i\pi\boldsymbol{\eta}\mathbf{f}} \tag{92}$$

and (84) shows that $\mathbf{T}$ is a unitary matrix.

The relationship of (88) and (92) is just that of (71) and (72). We suppose that the squared elements of $\mathbf{T}$ give the observable probabilities for the deflexion of incident particles from one momentum state to another in the course of the scattering process.

This procedure is actually equivalent to that of § 7 because, in problems to which both methods are applicable, the S matrix (88) is identical with the matrix S of (60).

10. CALCULATION OF T FROM S

Except in the first approximation (71) and (72) are not very convenient for the actual calculation of T from S. If in the block representation (64) we have

$$\eta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad S = \begin{pmatrix} a^\dagger & b^\dagger \\ c & d \end{pmatrix}, \quad S^* = \begin{pmatrix} a & -c^\dagger \\ -b & d^\dagger \end{pmatrix}, \tag{93}$$

the conditions $SS^* = S^*S = 1$ give

$$\left.\begin{aligned} a^\dagger a - b^\dagger b &= 1, & aa^\dagger - c^\dagger c &= 1, \\ dd^\dagger - cc^\dagger &= 1, & d^\dagger d - bb^\dagger &= 1, \\ ca &= db, & ab^\dagger &= c^\dagger d. \end{aligned}\right\} \tag{94}$$

The transformations $\begin{pmatrix} v \\ w \end{pmatrix} = S\begin{pmatrix} x \\ y \end{pmatrix}, \quad \begin{pmatrix} x \\ y \end{pmatrix} = S^*\begin{pmatrix} v \\ w \end{pmatrix},$

give

$$w = cx + dy, \quad x = av - c^\dagger w, \tag{95}$$

which imply

$$\begin{pmatrix} v \\ y \end{pmatrix} = T\begin{pmatrix} x \\ w \end{pmatrix},$$

with

$$T = \begin{pmatrix} a^{-1} & a^{-1}c^\dagger \\ -d^{-1}c & d^{-1} \end{pmatrix} = \begin{pmatrix} a^{-1} & a^{-1}c^\dagger \\ -ba^{-1} & d^{-1} \end{pmatrix}. \tag{96}$$

In the particular case of 2×2 matrices these results take the simple form

$$S = \begin{pmatrix} \bar{p} & \bar{q} \\ q & p \end{pmatrix} \Big/ r, \quad T = \begin{pmatrix} \bar{r} & \bar{q} \\ -q & r \end{pmatrix} \Big/ p, \tag{97}$$

with

$$|p|^2 - |q|^2 = |r|^2.$$

Let us see what corresponds to an 'almost certain' transition between two states of weights $\eta = +1$ and $\eta = -1$. Because of the normalization, in the S picture the system cannot pass from being wholly in a state in which $\eta = +1$ to being wholly in a state with $\eta = -1$. However, in the T picture the corresponding transition is not forbidden and (97) shows that the appropriate S and T matrices can be represented by

$$T = \begin{pmatrix} \epsilon & 1 \\ -1 & \epsilon \end{pmatrix}, \quad S = \begin{pmatrix} \dfrac{1}{\epsilon} & \dfrac{1}{\epsilon} - \tfrac{1}{2}\epsilon \\ \dfrac{1}{\epsilon} - \tfrac{1}{2}\epsilon & \dfrac{1}{\epsilon} \end{pmatrix}, \tag{98}$$

where ϵ is small.

11. SOME APPLICATIONS

In recent years various systems of Lorentz invariant linear wave equations have been proposed by different authors and a common feature of them all is that, as shown in A, for spin > 1 the formalism necessarily describes particles with more than one value of the rest mass. Blochinzev's discussion (1947) even allows a continuum of mass values. Interaction terms of the conventional type lead to transitions between the various mass states, so that attempts to restrict the particles to a single mass eigenvalue by the addition of subsidiary conditions lead to great difficulties. As stable particles of spin > 1 are not known in nature it may be doubted whether such restrictions are required. On the other hand, there is a wealth of experimental evidence for the existence of particles of various masses. In particular the recent experiments of Lattes, Occhialini & Powell (1947) at Bristol demonstrate the disintegration of heavy mesons into light mesons, the mass ratio being consistent with a theory of the type given by Bhabha in A. While it is not claimed that the Bristol experiments can be explained in terms of such formal theories, it is nevertheless desirable to study the transitions between the different mass states allowed by the wave equations.

Wild's paper implies quite generally that wave equations which give more than one mass value, cannot be quantized with a positive definite metric. It is a feature of all the simpler theories that if the mass eigenvalues χ/j are arranged in ascending order, the corresponding factors $D(r, j)$ alternate in sign. Now the signature of the η metric is derived from D by (51) or (55) and we see that, in such theories, the transition of a particle from a state of excited mass to the state of next lower mass is a transition between states of $\eta = +1$ and $\eta = -1$. The example (94) shows that the T formalism can be used to describe such transitions which are otherwise quite unintelligible.

The method of quantization and physical interpretation developed in this paper might also be used to avoid the negative energies which appear in the generalized electrodynamics of Podolsky & Schwed (1948).

12. TRANSITIONS BETWEEN MASS STATES

(*a*) *The two-level problem*

Consider a transition of a particle between neighbouring mass states by emission of a quantum of radiation. We have conservation of E_0, $\mathbf{G}$, Q and it is obvious from (52) or (58) that both initial and final states must have the same sign of k_0, say $k_0 > 0$. We are concerned with states

$$n'_a = 1, n'_b = 0, \ldots, n'(\lambda) = 0 \quad \text{and} \quad n'_a = 0, n'_b = 1, \ldots, n'(\lambda) = 1$$

of the complete system, these we shall denote by $(a, 0 \,|$ and $(b, \lambda \,|$ and we shall write E_a and E_b for $k_0(a)$ and $k_0(b)$.

We suppose $E_a > E_b$, and have the conservation laws

$$\mathbf{k}(a) = \mathbf{k}(b) + \boldsymbol{\lambda}, \tag{99}$$

$$E_a - E_b \doteq \lambda_0. \tag{100}$$

It is convenient to choose the frame of reference so that $\mathbf{k}(a) = \mathbf{0}$. Only one initial state $(a, 0)$ is involved, but because of the degeneracy of the radiation field there are many final states (b, λ) satisfying (99) and (100).

In the case of a positive definite metric the problem has been solved by Weisskopf & Wigner (see Heitler 1944, § 12) and their solution applies to our problem if η_a and η_b have the same sign. We are interested here in the case in which η_a and η_b are of different sign so that $\eta_a \eta_b = -1$.

It is sufficient to consider only those rows and columns in which S and T differ from the unit matrix so that, in the block representation (93) the labelling of the rows is

$$\begin{pmatrix} a, 0 \\ b, \lambda \end{pmatrix},$$

and so the leading submatrix of S reduces to a single number. The transformation (96) then shows that if the leading diagonal element of T is $e^{-\frac{1}{2}\gamma t}$ that of S is $e^{\frac{1}{2}\gamma t}$. As usual γ^{-1} is to be interpreted as the lifetime of the material system.

The calculation of γ follows that given in Heitler's book (1944). The equations are

$$\left.\begin{aligned} i\frac{d}{dt}(a,0\,|\,t) &= \sum_{\lambda}(a,0\,|\,H\,|\,b,\lambda)\,(b,\lambda\,|\,t)\exp\left[-i(E_a-E_b-\lambda_0)\,t\right],\\ i\frac{d}{dt}(b,\lambda\,|\,t) &= (b,\lambda\,|\,H\,|\,a,0)\,(a,0\,|\,t)\exp\left[i(E_a-E_b-\lambda_0)\,t\right], \end{aligned}\right\} \tag{101}$$

where the $(a,0\,|\,H\,|\,b,\lambda)$ are the time independent matrix elements of the interaction Hamiltonian.

The initial conditions are

$$(a,0\,|\,t=0)=1,\quad (b,\lambda\,|\,t=0)=0,$$

and we assume

$$(a,0\,|\,t)=e^{\frac{1}{2}\gamma t}. \tag{102}$$

We obtain

$$(b,\lambda\,|\,t) = -(b\lambda\,|\,H\,|\,a,0)\frac{\exp\left[i(\nu_0-\lambda_0)\,t+\frac{1}{2}\gamma t\right]-1}{\nu_0-\lambda_0-\frac{1}{2}i\gamma}, \tag{103}$$

where

$$\nu_0 = E_a-E_b. \tag{104}$$

Now $H=H^*$ implies

$$(a,0\,|\,H\,|\,b,\lambda)\,(b,\lambda\,|\,H\,|\,a,0)=\eta_a\eta_b\,|\,(a,0\,|\,H\,|\,b,\lambda)\,|^2,$$

so that, substituting from (103) to (101), we obtain

$$-\tfrac{1}{2}i\gamma = \eta_a\eta_b\sum_{\lambda}|\,(a,0\,|\,H\,|\,b,\lambda)\,|^2\left\{\frac{1-\exp\left[i(\lambda_0-\nu_0)\,t-\frac{1}{2}\gamma t\right]}{\nu_0-\lambda_0-\frac{1}{2}i\gamma}\right\}. \tag{105}$$

Since $\eta_a\eta_b=-1$, this is the same as Heitler's equation (8) except that on the right-hand side the sign of γ is reversed. This does not affect the summation over the oscillators λ so that we obtain Heitler's equation (11)

$$\gamma = 2\pi\rho\int|\,(a,0\,|\,H\,|\,b,\lambda)\,|^2\,d\Omega. \tag{106}$$

Thus we see that in our formalism the calculation of the lifetime from the Hamiltonian is the same whether or not we are working in an indefinite metric.

This problem involves only one particle so that equations (101) are the same as the equations of the c number theory. In that theory we should therefore have the solutions (102) and (103), which imply that for large t the a degree of freedom contains positive energy and charge proportional to $e^{\frac{1}{2}\gamma t}$ and that this is compensated by the appearance of negative energy and charge in the b degree of freedom and by the emission of radiation. Such results are obviously unsuitable for direct physical interpretation.

(b) *Application to Bhabha's theory*

As an example consider the theory classified by Bhabha as $(\frac{3}{2},\frac{1}{2})$. In this case if the mass in the ground state is M that in the excited state is $3M$. The conservation laws (99) and (100) show that, in the frame of reference of zero total momentum, the energy of the emitted quantum is $k=\frac{4}{3}Mc^2$. If the free particle wave function in state a is $\phi_a\,e^{-i(kx)}$ the matrix element for the transition is

$$(a,0\,|\,H\,|\,b,\lambda)=\{-e\sqrt{(2\pi\hbar^2c^2/k)}\}\,\phi_b^{\dagger}D\alpha_l\phi_a, \tag{107}$$

where α_l is the component of α in the direction of polarization of the photon. The important factor is that enclosed in curly brackets; the other has been evaluated but, for the purpose of this paper, it is sufficient to remark that it is of order unity. If for M we insert the mass of a proton and put $e^2/\hbar c = \frac{1}{137}$ we find that the lifetime γ^{-1} of the excited state is approximately 10^{-22} sec. This is much too short to allow direct observation of the excited state though the emitted photon might be observable.

12. DISCUSSION

We have shown that our theory can always be quantized in a consistent way which deals only with positive numbers of particles. In the case of Bose statistics all particles have positive energy but either sign of the charge; in the case of Fermi statistics we have formally positive definite charge and indefinite energy but, by the familiar hole interpretation, a description in terms of particles of positive energy is possible. Consistently with all conservation laws, we have described transition probabilities in terms of a unitary matrix T so that all requirements of the theory of probability are satisfied. In the relativistic region this is all that can be asked of any theory. On the other hand, in the non-relativistic limit we are required to give an account of closed stationary states. It must be assumed, as a restriction on the α matrices, that in the non-relativistic limit only states with $\eta = +1$ are excited and then the ordinary methods of the positive definite metric are available; in fact the S and T matrices become identical. In the non-relativistic limit we are concerned only with the states of lowest rest mass, and so we require that, in the c number theory, all such states have the same sign of the energy and of the charge. This is equivalent to the condition that, if j is the greatest positive eigenvalue of α_0, $D(r, j) = 1$ for all r and this requirement is easily satisfied.

Our discussion has not distinguished between integral and half odd integral spins. The cases of spin 0, 1 and of spin $\frac{1}{2}$ are special in that they have respectively $D = 1$, $s = \pm 1$ (for $j \neq 0$) and $s = 1, D = \pm 1$ and it is well known that quantization with a positive definite metric is possible if we choose Bose statistics for spins 0, 1 and Fermi statistics for spin $\frac{1}{2}$. But in all generalized wave equations both s and D are of indefinite sign and no such special considerations exist. Quantization of such equations is possible only by overcoming the difficulties (Pauli 1940) which arise when we try to quantize half odd spins with Bose statistics or integral spins with Fermi statistics. This we have done and so, in general, our formalism implies no connexion between spin and statistics.

Our procedure enables us to answer all questions which, according to the point of view of Heisenberg & Møller, are of any physical significance. By avoiding the requirement that in the c number theory either the charge or the energy must be positive definite, it removes a very severe restriction on the possible relativistic wave equations and, for example, makes possible the description of particles which can exist in several different mass and spin states. On the other hand, in the sort of description of Nature which we envisage, there is another important requirement, namely that all states must be specified by dynamical variables of real physical significance. For example we may describe the state of a particle by its charge, momentum and mass and by its spin in the frame of reference in which it is at rest. This description is unambiguous only if no spin representation occurs more than once with each mass value. We require therefore that none of the irreducible representations of the three-dimensional rotation group which can be obtained by reduction of the spin matrices I_{23}, I_{31}, I_{12}, introduced in (2), occurs more than once with each eigenvalue of α_0. This is quite a severe restriction, for example in Bhabha's theory in which the possible wave equations are classified as (n, m) it requires $m = 0$ or $m = \frac{1}{2}$ or $m = n$ (see the appendix).

We saw in § 8 that the wave function $|\,0)$ which forms the initial data for the solution of the Schrödinger equation (59) refers to physical conditions partly in the observer's past and partly in his future. We have here a situation rather similar to that arising in Dirac's (1938) classical theory of the electron in which 'we must obtain solutions of the equations of motion for which the initial position and velocity of the electron are prescribed, together with its final acceleration, instead of solutions with all the initial conditions prescribed'. This similarity is probably not a coincidence. For suppose that we have a wave equation for a particle which can interact with another field and that we eliminate the effects of the field created by the particle in terms of the dynamical variables describing the particle itself. We have then an equation for a particle which interacts only with external fields. This new equation may be expected to involve, in addition to the ground state, excited states in which the particle appears with higher mass and spin values. In the meson theory such excited levels have been calculated by means of the strong coupling approximation. It is inevitable that the new equation will be more complicated than the original and so, if it can be expressed as a linear wave equation, must correspond to something other than the simple equations of spin 0, $\frac{1}{2}$, 1. Now Dirac in his classical theory of electrons carries through such an elimination so that it is not surprising that his equations should show this apparent interconnexion between past and future which, as we have seen, is typical of all the more general relativistic wave equations.

Finally I should like to thank Professor Rosenfeld for his interest in this work.

APPENDIX

The Lorentz group in five *dimensions*

In Bhabha's notation (A), we consider the irreducible representation $R_5(n, m)$ of the five-dimensional Lorentz group. The matrices are

$$I^{40};\quad I^{14}, I^{24}, I^{34};\quad I^{10}, I^{20}, I^{30};\quad I^{23}, I^{31}, I^{12}. \tag{108}$$

I^{40} commutes with the representation I^{23}, I^{31}, I^{12} of the three-dimensional rotation group and we proceed to determine how the various irreducible parts D_K are associated with the different eigenvalues of I^{40}.

The matrices $$I^{20}, I^{30}, I^{40};\quad I^{34}, I^{42}, I^{23}, \tag{109}$$

define the reducible representation

$$\left.\begin{array}{l} \left(\frac{n+m}{2},\frac{n-m}{2}\right)+\left(\frac{n+m-1}{2},\frac{n-m+1}{2}\right)+\ldots+\left(\frac{n-m}{2},\frac{n+m}{2}\right), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad\qquad\qquad \ldots\ldots\ldots\ldots\ldots\ldots, \\ (m,0) \quad + \qquad\qquad \ldots \qquad\qquad \ldots+ \quad (0,m), \end{array}\right\} \tag{110}$$

of the proper four-dimensional Lorentz group. In the representation $\left(\frac{n+m}{2},\frac{n-m}{2}\right)$ the combinations

$$\left.\begin{array}{lll} \tfrac{1}{2}(I^{40}+iI^{23}), & \tfrac{1}{2}(I^{30}+iI^{42}), & \tfrac{1}{2}(I^{20}+iI^{34}), \\ \tfrac{1}{2}(I^{40}-iI^{23}), & \tfrac{1}{2}(I^{30}-iI^{42}), & \tfrac{1}{2}(I^{20}-iI^{34}), \end{array}\right\} \tag{111}$$

behave as independent three-dimensional angular momenta of magnitudes $\frac{1}{2}(n+m)$ and $\frac{1}{2}(n-m)$ respectively. Hence by the rules for combination of angular momenta we may write down the possible eigenvalue combinations of I^{40} and iI^{23}. We obtain all the eigenvalues of I^{23} associated with a particular eigenvalue of I^{40} by combining the contributions from all the representations in the direct sum (111). It is then a straightforward jig-saw puzzle to arrange these eigenvalues of I^{23} by representations of the three-dimensional rotation group.

We find that the representation of the rotation group corresponding to the eigenvalue s of I^{40} is

$$D_m\times\{D_{n-s}+D_{n-s-2}+\ldots+(D_1 \text{ or } D_0)\} \quad (\text{for } n\geqslant s\geqslant m), \tag{112}$$

$$\left.\begin{array}{l} D_n+D_{n-1}+\ldots+D_{s+n-m+1} \\ +D_{n-1}+D_{n-2}+\ldots+D_{s+n-m}+D_s\times\{D_{n-m}+D_{n-m-2}+\ldots+(D_1 \text{ or } D_0)\} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ +D_m+D_{m-1}+\ldots+D_{s+1} \quad (\text{for } m-1\geqslant s\geqslant 0). \end{array}\right\} \tag{113}$$

We note that in the special case $n = m$ the result is very simple, we have

$$D_n+D_{n-1}+\ldots+D_s. \tag{114}$$

Now in Bhabha's wave equation $\alpha^0 = I^{04}$ so that, by use of (112) and (113) we may obtain the result quoted in § 13.

REFERENCES

Bhabha, H. J. 1945 *Proc. Indian Acad. Sci.* A, **21**, 241.
Bhabha, H. J. 1947 *Rep.* 1946 *Cambridge Conf.* **1**, 22. (Referred to as A.)
Blochinzev, D. 1947 *J. Phys., U.S.S.R.*, **11**, 72.
Chang, T. S. 1947 *Proc. Camb. Phil. Soc.* **43**, 183.
Dirac, P. A. M. 1938 *Proc. Roy. Soc.* A, **167**, 148.
Dirac, P. A. M. 1942 *Proc. Roy. Soc.* A, **180**, 1.
Dirac, P. A. M. 1943 *Comm. Dublin. Inst. Adv. Studies*, Ser. A, no. 1.
Dirac, P. A. M. 1947 *Quantum mechanics*, 3rd ed. Oxford:
Heisenberg, W. 1943 *Z. Phys., Lpz.*, **120**, 513.
Heitler, W. 1944 *The quantum theory of radiation*, 2nd ed. Oxford Univ. Press.
Lattes, C. M., Occhialini, G. P. S. & Powell, C. F. 1947 *Nature*, **160**, 453 and 486.
Le Couteur, K. J. 1948 *Proc. Camb. Phil. Soc.* **44**, 63. (Referred to as B.)
Møller, C. 1947 *Rep.* 1946 *Cambridge Conf.* **1**, 194.

Pauli, W. 1940 *Phys. Rev.* **58**, 716.
Pauli, W. 1941 *Rev. Mod. Phys.* **13**, 203.
Pauli, W. 1943 *Rev. Mod. Phys.* **15**, 175.
Pauli, W. 1947 *Rep.* 1946 *Cambridge Conf.* **1**, 5.
Podolsky, B. & Schwed, P. 1948 *Rev. Mod. Phys.* **20**, 40.
Weiss, P. 1938 *Proc. Roy. Soc.* A, **169**, 102.
Wentzel, G. 1947 *Rev. Mod. Phys.* **19**, 1.
Wild, E. 1947 *Proc. Roy. Soc.* A, **191**, 253.

The Gouy diffusiometer*; further calibration

BY A. G. OGSTON, *Department of Biochemistry, University of Oxford*

(*Communicated by R. A. Peters, F.R.S.—Received* 22 *July* 1948)

The application of the modified ray theory given by Kegeles & Gosting (1947) to the Gouy diffusiometer (Coulson, Cox, Ogston & Philpot 1948) has been tested. It has been shown that experimental results, both with single homogeneous substances and with mixtures, agree well with its predictions. A simple method for analyzing the diffusion data given by a mixture of two diffusing substances has been worked out and has been applied to the correction of the apparent diffusion constant of lactoglobulin contaminated by some heterogeneous material and to known mixtures of lactoglobulin with sucrose and potassium chloride. The necessary computations have been performed over a useful range of the variables and the results tabulated. Certain improvements in the experimental technique are described.

INTRODUCTION

In the description of an interferometric method of measuring diffusion constants given by Coulson, Cox, Ogston & Philpot (1948), the movement of the outermost dark band was shown to be described adequately by a mathematical approximation using Bessel functions, for values of the total refractive difference between solution and solvent greater than about 13 wave-lengths' path difference. With smaller differences of refractive index two sources of inaccuracy appeared:

(*a*) The factor β was not accurately given by our expression.

(*b*) The apparent movement of the outermost band became periodic, owing to the crossing of the moving interference bands due to the diffusion boundary and the fixed Fraunhofer bands arising from the rectangular stop; the intensity of the latter was insufficient to cause the appearance of this effect with greater differences of refractive index. The Bessel function approximation was inadequate for the description of positions of bands other than the outermost and this prevented us from using the relative positions of the inner bands to test the departure of the boundary from the ideal Gaussian form.

Since our work went to press, independent work by Kegeles & Gosting (1947) and Longsworth (1947) on the same method of observing diffusion has appeared. These authors have observed the interference pattern at times much longer than those used by us: they have shown, at these longer times, that the movement of all but the innermost bands of the pattern is satisfactorily described by a modification of the ray-theory treatment, by which the intensity minima are taken to occur at phase-differences of $j+\frac{3}{4}$ instead of $j+\frac{1}{2}$ (j being an integer). This discovery makes possible the accurate description of the movement of the whole band system and so leads to the possibility of testing the form of the boundary; it also makes possible the use of the method at lower differences of refractive index between solution and solvent and thus the use of a calibrating substance under these conditions is avoided.

---

* We have agreed with our American colleagues to call this general method 'The Gouy interference method for diffusiometry' and we propose to call our version of the method 'The Gouy diffusiometer'.

Experiments and computations are described here which extend the useful range of our version of the method down to a phase-difference of a little more than one wave-length and which make possible some degree of analysis of the homogeneity of a diffusing substance.

THEORY

(1) *Diffusion of a single substance*

(*a*) For a simple Gaussian boundary, between solution and solvent whose refractive indices differ by v wave-lengths (taking into consideration the length of the cell) the positions of intensity minima are described by

$$j+\tfrac{3}{4} = v\left\{\frac{2}{\sqrt{\pi}}\int_0^z e^{-z^2}\,dz - \frac{2z\,e^{-z^2}}{\sqrt{\pi}}\right\}, \tag{1}$$

where z is related to distance x from the centre of the boundary and to the time t since the beginning of diffusion by $z = x/\sqrt{(4Dt)}$.

Following Kegeles & Gosting (1947), the expression in the bracket will be written as $f(z)$.

The angular deflexion θ_j of any one of these minima, divided by the maximum theoretical deflexion in the middle of the boundary ($z=0$) $\theta_{\text{max.}}$ is given by

$$\theta_j/\theta_{\text{max.}} = e^{-z^2}. \tag{2}$$

The relative positions of the minima, from the optic axis, are thus independent of time and of v. It is convenient also to express the values of $j+\frac{3}{4}$ so as to obtain a function independent of v. This can be done by giving a number $n_0 = v$ ($j=0$) to the outermost minimum and numbers $n_j = (n_0 - j)$ to the inner minima. Then

$$\frac{n_j-\frac{3}{4}}{n_0} = 1-\frac{j+\frac{3}{4}}{v} = 1-f(z). \tag{3}$$

Thus, whatever the values of v and of t, a given value of $\dfrac{n_j-\frac{3}{4}}{n_0}$ should correspond with a certain value of $\theta_j/\theta_{\text{max.}}$. It has been found most convenient to plot $\log\left(\dfrac{n_0}{n_j-\frac{3}{4}}\right)$ against $\log\theta_{\text{max.}}/\theta_j$; departure from this relationship shows departure of the boundary distribution from the simple Gaussian form.

(*b*) The value of β for the outermost minimum is given by $\theta_0/\theta_{\text{max.}}$; β obtained from this treatment is written β_T to distinguish it from the value obtained from Bessel functions, β_B.

(*c*) The combination between the interference patterns arising from the boundary and from the stop, which causes the periodic movement of the bands at low values of v, could be avoided by the use of a very much wider stop as was in effect done by Longsworth (1947). On the other hand, the small (2 mm.) stop is useful in reducing the intensity of light in the innermost bands, which arise from the outer parts of the boundary: this avoids the photographic difficulties mentioned by Longsworth by bringing the intensity of the whole pattern within the latitude of the photographic plate. The theoretical positions of zero intensity in the combined pattern can be defined, but it is doubtful how closely these positions would correspond with the apparent centres of the dark bands obtained in the photographic record, since the latter would depend on the distribution of light intensity on each side of the theoretical zero and on the characteristics of the photographic plate. It seems likely, however, that the theoretical zero and the apparent centre of the band will correspond fairly closely where a minimum of the boundary pattern coincides with a Fraunhofer minimum, since in this case the distribution of light about the minimum is fairly symmetrical.

(2) *Diffusion of a mixture*

(*a*) *The form of the interference pattern.* With a mixture of substances, which diffuse independently of each other from a common initially sharp boundary, the distribution of each component is described by a Gaussian curve, all curves having the same origin. Assuming that the minima of intensity still occur at phase-differences of $j+\frac{3}{4}$, equations (1) and (2) become

$$j+\tfrac{3}{4} = \Sigma v_i f(z_i), \tag{4}$$

and

$$\theta_j/\theta_{\text{max.}} = \frac{\Sigma\{v_i e^{-z_i^2}/\sqrt{D_i}\}}{\Sigma\{v_i/\sqrt{D_i}\}}, \tag{5}$$

where subscript i refers to the ith component of the mixture. Both expressions are independent of time. Individual values of v_i may be positive or negative and these cause the values of $\log \theta_{\text{max.}}/\theta_j$ for a given value of $\log \dfrac{n_0}{n_j - \frac{3}{4}}$ to be greater or less, respectively, than if there were only one diffusing component.

(*b*) *Method of analysis.* It would be possible, in principle, to solve equations (4) and (5) for as many values of v_i and of D_i as there are bands in the interference pattern, by measuring the values of θ_j and n_j. Since no simple way of performing this solution presented itself, it was necessary to proceed by computation. Attention has been confined to the case of two components only since it is likely that analysis in these terms will nearly enough represent the behaviour of preparations which are not quite homogeneous and will enable a good approximation to the diffusion constant of the main component to be obtained. It is unlikely that the resolving power of the method is good enough to justify analysis in terms of several components or that, in deliberate work on mixtures, more than two diffusing components will be used.

(*c*) *Values of $\bar{\beta}$ and of $\bar{D}$ of a 2-component mixture.* If $\bar{\beta} = \theta_0/\theta_{\text{max.}}$ is known, then $\bar{D}$ for the mixture is obtained as in the case of one component and is related to D_1 and D_2 of the two components by

$$\frac{v_1+v_2}{\sqrt{\bar{D}}} = \frac{v_1}{\sqrt{D_1}} + \frac{v_2}{\sqrt{D_2}} \quad \text{or} \quad \sqrt{\frac{D_1}{\bar{D}}} = \alpha + (1-\alpha)\sqrt{\frac{D_1}{D_2}}, \tag{6}$$

where $\alpha = v_1/v_1+v_2$. However, $\bar{\beta}$ is less than the value of β_T for a single component having the same total refractive difference. From equations (4) and (5), it is given by

$$\bar{\beta} = \theta_0/\theta_{\text{max.}} = \frac{\{\alpha/(1-\alpha)\,a\}\,e^{-z_1^2} + e^{-a^2z_1^2}}{\{\alpha/(1-\alpha)\,a\}+1}, \tag{7}$$

where $a = \sqrt{(D_1/D_2)}$ and z_1 is the solution of

$$\frac{n_0-\frac{3}{4}}{n_0} = 1 - \{\alpha f(z_1) + (1-\alpha)\,f(az_1)\}. \tag{8}$$

COMPUTATION

(*a*) *One diffusing component.* Equations (2) and (3) have been evaluated for a range of values of z. The resulting values of $\log n_0/(n_j - \frac{3}{4})$ and $\log \theta_{\text{max.}}/\theta_j$ are given in table 1 and figure 1. From the same computation, the values of $\beta_T = \theta_0/\theta_{\text{max.}}$ for

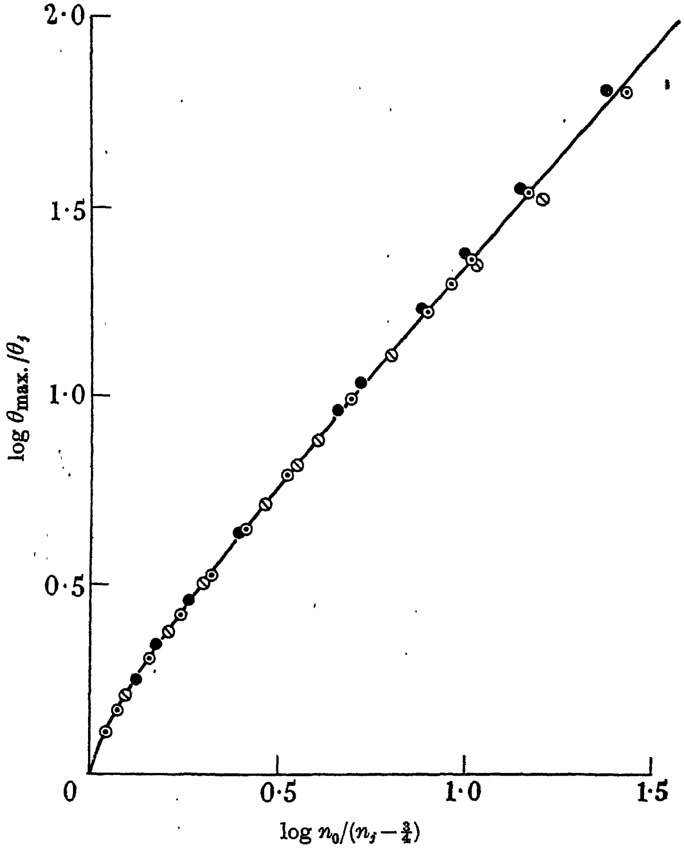


FIGURE 1. Full line: theoretical curve (table 1). Open circles: experimental values for glycine. Filled and barred circles: experimental values for two samples of lactoglobulin.

different values of $n_0 = v$ were obtained and are given, together with the corresponding values of β_B, in table 2. Interpolation is conveniently performed from a plot of β_T against $1/v$.

TABLE 1

$\log n_0/(n_j-\frac{3}{4})$	$\log \theta_{\max.}/\theta_j$	$\log n_0/(n_j-\frac{3}{4})$	$\log \theta_{\max.}/\theta_j$
0	0	0·386	0·626
0·008	0·039	0·473	0·734
0·019	0·070	0·676	0·977
0·037	0·109	0·912	1·255
0·061	0·152	1·043	1·407
0·134	0·278	1·335	1·737
0·242	0·434	1·664	2·105

TABLE 2

v	β_T	β_B	v	β_T	β_B
∞	1	1	3·87	0·613	0·593
39·06	0·913	0·913	2·83	0·528	0·497
17·20	0·852	0·850	1·75	0·368	0·307
9·24	0·779	0·772	1·27	0·237	0·142
5·70	0·697	0·686	1·13	0·1844	0·073

(*b*) *Two diffusing components.* It has been found most convenient in practice to make use of the differences between the values of $\log \theta_{\max.}/\theta_j$ for one and two components at two stated values of $\log n_0/(n_j-\frac{3}{4})$, since these values can be easily determined from the experimental record. These differences are uniquely determined by the values of $a = \sqrt{(D_1/D_2)}$ and of $\alpha = v_1/(v_1+v_2)$. 0·2 and 0·9 were chosen as the stated values of $\log n_0/(n_j-\frac{3}{4})$; the corresponding one-component values of $\log \theta_{\max.}/\theta_j$ are 0·3749 and 1·2426.

Equation (4) was used in the form

$$\log \frac{n_0}{n_j-\frac{3}{4}} = -\log\left[1-\{\alpha f(z_1)+(1-\alpha)f(az_1)\}\right].$$

For each chosen value of a and α, values of z_1 were chosen such as to make the expression have values a little less than and a little greater than its stated value (0·2 or 0·9); the same values of z_1 were used to calculate $\log \theta_{\max.}/\theta_j$, from equation (5) put into the form

$$\log \theta_{\max.}/\theta_j = \log\left\{\frac{\alpha}{(1-\alpha)a}+1\Big/\frac{a}{(1-\alpha)a}e^{-z_1^2}+e^{-(az_1)^2}\right\}.$$

The value of $\log \theta_{\max.}/\theta_j$ corresponding to the stated value of $\log n_0/(n_j-\frac{3}{4})$ was then obtained with sufficient accuracy by linear interpolation. The difference between such a value of $\log \theta_{\max.}/\theta_j$ and its one-component value is called $\Delta_{0\cdot 2}$ or $\Delta_{0\cdot 9}$. These computations were performed over most of the range $a^2 = 0$ to 0·5 and $\alpha = 1$ to 0·05.

The results are given in tables 3 and 4; table 3 shows the values of $\Delta_{0\cdot 9}$ which correspond with stated (interpolated) values of $\Delta_{0\cdot 2}$ and of a^2; table 4 shows the values of $\Delta_{0\cdot 9}$ which correspond with stated values of $\Delta_{0\cdot 2}$ and of α, the values in heavy type being limits corresponding with $a^2 = 0$. For use, these values are best plotted as lines of constant a^2 and α respectively on graphs of $\Delta_{0\cdot 2}$ against $\Delta_{0\cdot 9}$.

TABLE 3. VALUES OF $\Delta_{0\cdot9}$ FOR GIVEN VALUES OF $\Delta_{0\cdot2}$ AND OF a^2

$\Delta_{0\cdot2}$ \ a^2	0	0·01	0·025	0·05	0·10	0·15	0·20	0·25	0·30	0·40	0·50
0·0015	0·0158	0·0152	0·0145	0·0135	0·0122	0·0110	0·0101	0·0095	0·0090	0·0081	0·0074
0·003	— 0·0307	— 0·0298	— 0·0285	— 0·0268	0·001 0·0245	0·002 0·0225	0·0025 0·0210	0·003 0·0192	0·0035 0·0180	0·0045 0·0160	0·006 0·0142
0·005	— 0·052	— 0·051	— 0·048	0·002 0·045	0·003 0·041	0·004 0·038	0·005 0·035	0·006 0·032	0·007 0·030	0·009 0·026	0·012 0·022
0·007	— 0·074	— 0·071	— 0·067	0·003 0·062	0·005 0·057	0·0065 0·052	0·008 0·048	0·010 0·044	0·011 0·040	0·013 0·034	0·017 0·028
0·010	— 0·110	— 0·105	— 0·098	0·005 0·091	0·008 0·080	0·010 0·072	0·013 0·065	0·015 0·060	0·017 0·055	0·020 0·045	0·027 0·035
0·012	— 0·136	— 0·126	— 0·117	— 0·109	0·010 0·092	0·013 0·086	0·016 0·078	0·018 0·070	0·020 0·064	0·025 0·050	0·036
0·015	— 0·172	— 0·160	— 0·150	— 0·138	0·013 0·120	0·017 0·105	0·020 0·093	0·023 0·082	0·026 0·074	0·031 0·057	
0·020	— 0·239	— 0·220	— 0·205	— 0·186	0·019 0·159	0·024 0·135	0·028 0·116	0·032 0·102	0·036 0·089	0·050	
0·025	— 0·323	— 0·290	— 0·266	— 0·238	0·025 0·198	0·031 0·161	0·036 0·138	0·041 0·120	0·047 0·102		
0·030	— 0·420	— 0·364	— 0·327	— 0·286	0·032 0·228	0·038 0·186	0·044 0·151	0·051 0·128	0·060 0·105		
0·035	— 0·575	— 0·420	— 0·385	— 0·335	0·041 0·255	0·046 0·205	0·052 0·165	0·061 0·130			
0·040	— —	— —	— —	— 0·365	0·046 0·268	0·054 0·220	0·062 0·175	0·074 0·125			
0·045	— —	— —	— —	— 0·390	0·053 0·280	0·063 0·230	0·074 0·175	0·100			
0·06	— —	— —	— —	— —	0·070 0·290	0·086 0·226	0·114 0·152				
0·08	— —	— —	— —	— —	0·098 0·290	0·134 0·184					
0·10	— —	— —	— —	— —	0·126 0·265						
0·12	— —	— —	— —	— —	0·167 0·224						

TABLE 4. VALUES OF $\Delta_{0\cdot9}$ FOR GIVEN VALUES OF $\Delta_{0\cdot2}$ AND OF α

$\Delta_{0\cdot2}$ \ α	0·99	0·98	0·95	0·90	0·85	0·80	0·70	0·60	0·50	0·40	0·30	0·20	0·10	0·05
0·001	0·0060	0·0055	0·0050	0·0046										
0·002	0·0170	0·0125	0·0105	0·0090	0·0083									
0·0036	**0·037**	0·029	0·021	0·017	0·015	0·014								
0·005	—	0·046	0·032	0·026	0·022	0·020	0·018	0·017	0·015	0·014	0·013	0·012	0·011	0·009
0·0073	—	**0·078**	0·055	0·041	0·035	0·030	0·027	0·025	0·023	0·021	0·019	0·017	0·015	0·012
0·010	—	—	0·085	0·060	0·050	0·045	0·040	0·036	0·032	0·028	0·025	0·022	0·019	0·015
0·013	—	—	0·122	0·086	0·071	0·061	0·053	0·046	0·042	0·038	0·033	0·028	0·023	0·018
0·015	—	—	0·153	0·105	0·086	0·073	0·061	0·053	0·048	0·042	0·037	0·032	0·025	0·021
0·0193	—	—	**0·227**	0·153	0·119	0·100	0·079	0·069	0·062	0·054	0·047	0·040	0·031	0·025
0·025	—	—	—	0·237	0·175	0·139	0·109	0·089	0·081	0·069	0·058	0·049	0·037	0·031
0·030	—	—	—	0·335	0·230	0·179	0·135	0·107	0·095	0·080	0·065	0·055	0·043	0·037
0·035	—	—	—	0·464	0·292	0·220	0·165	0·124	0·107	0·090	0·075	0·062	0·048	
0·0394	—	—	—	**0·724**	0·360	0·258	0·191	0·140	0·118	0·099	0·082	0·068	0·052	
0·045	—	—	—	—	—	0·315	0·222	0·161	0·131	0·109	0·090	0·074	0·058	
0·050	—	—	—	—	—	0·364	0·252	0·179	0·143	0·117	0·097	0·080	0·064	
0·060	—	—	—	—	—	0·477	—	0·217	0·166	0·134	0·110	0·091	0·073	
0·070	—	—	—	—	—	—	—	—	0·188	0·150	0·124	0·102		
0·080	—	—	—	—	—	—	—	—	0·210	0·168	0·138	0·113		
0·100	—	—	—	—	—	—	—	—	0·251	0·200	0·163	0·130		
0·120	—	—	—	—	—	—	—	—	—	0·228	0·183			

By a similar process, from equations (7) and (8), the values of $\bar{\beta}$ for a range of a^2 and of α and for $v_1+v_2 = 30$, 21·4 and 15 have been computed and are given in table 5; these may be plotted as lines of constant a on a graph of $\bar{\beta}/\beta_T$ against α.

A few representative values of Δ were also computed for negative values of v_2. Their values were negative and numerically closely similar to those obtained for the same positive values of $v_2/(v_1+v_2)$ and the tables can therefore be used for interpreting negative as well as positive values of Δ by making a change in the sign of v_2.

TABLE 5. VALUES OF $\bar{\beta}/\beta_T$ FOR THE OUTERMOST BAND, FOR GIVEN VALUES OF a^2 α, AND OF v_1+v_2

a^2	v_1+v_2	α					
		0·9	0·8	0·7	0·5	0·3	0·1
0·5625	30	1·000	0·999	0·998	0·998	0·998	0·998
	21·4	1·000	0·999	0·998	0·996	0·997	0·998
	15	1·000	0·998	0·997	0·994	0·995	0·998
0·25	30	0·999	0·997	0·996	0·989	0·988	0·992
	21·4	0·998	0·994	0·993	0·986	0·984	0·990
	15	0·997	0·992	0·989	0·983	0·979	0·988
0·16	30	0·998	0·994	0·991	0·983	0·977	0·984
	21·4	0·995	0·991	0·987	0·978	0·972	0·979
	15	0·991	0·985	0·982	0·971	0·961	0·971
0·0625	30	0·996	0·990	0·984	0·969	0·954	0·950
	21·4	0·992	0·985	0·978	0·960	0·942	0·938
	15	0·987	0·978	0·971	0·949	0·924	0·921
0·01	30	0·994	0·986	0·975	0·950	0·910	0·849
	21·4	0·991	0·980	0·966	0·936	0·887	0·810
	15	0·985	0·970	0·957	0·918	0·852	0·756
0	30	0·993	0·983	0·969	0·934	0·859	0·608
	21·4	0·991	0·976	0·958	0·915	0·822	0·507
	15	0·980	0·964	0·948	0·890	0·775	0·366

$\bar{\beta}/\beta_T = 1{\cdot}000$ when $\alpha = 1$, for all finite values of a^2; and when $a^2 = 1$, for all finite values of α.

EXPERIMENTAL

(1) *Apparatus and its use*

The method used was that described by Coulson *et al.* (1948), with the following modifications:

(*a*) A new means of stopping the outflow from the cell has been devized (figure 2), because the method of freezing the outlet tip proved to be inconvenient, especially in experiments of longer duration, and because it occasionally caused disturbance of the boundary.

On the side of the regulating tap *A* distant from the cell, the capillary outlet tube is bent downward, first at 45° and then vertically, to end in a sharply tapered tip. A short open capillary side-arm *B* is sealed to the 45° portion and is fitted with a sac *C* made of 2 mm. bore flexible tubing (preferably of PVC since it is more rigid than rubber). The sac is enclosed in a hole in a Perspex block *D*, which is cemented on to the outlet tube and which carries a screw *E* which bears on the sac. The sac is filled with clean mercury. When it is desired to stop the outflow, the sac is compressed by means of the screw; a small drop of mercury is squeezed into the outlet tube, is carried down towards the outlet and there stops in the tapered tip, arresting the outflow smoothly and completely. A single filling of the sac is sufficient for many runs. The mercury droplet is removed from the tip, when required, by applying suction through a rubber tube. This system is completely positive and has the special advantage that there is an interval, during which outflow continues, between the turning of the screw and the arrival of the droplet at the tip, so that the effect of any mechanical disturbance produced in turning the screw is removed before the flow stops.

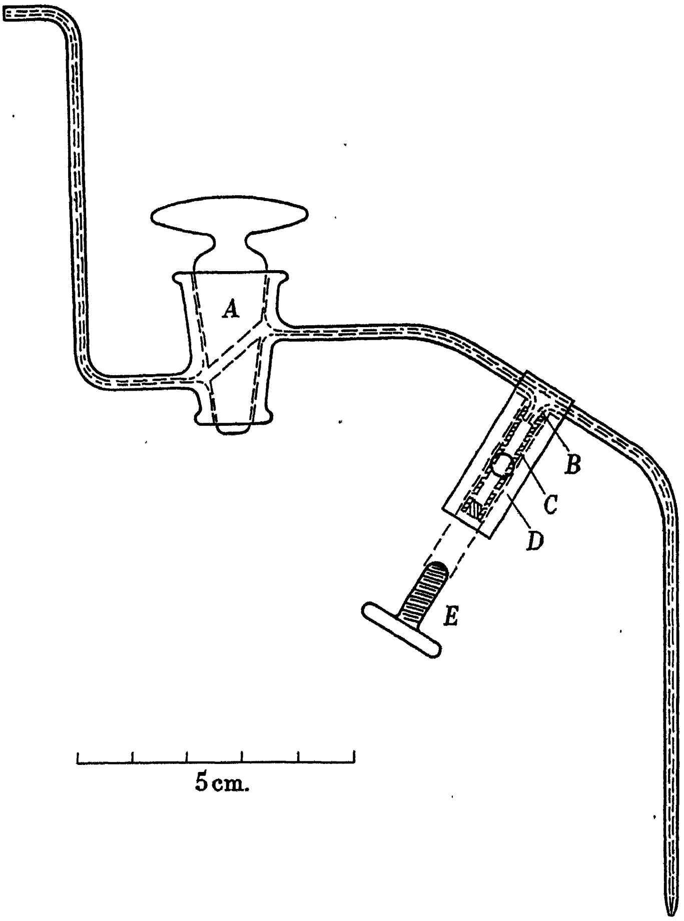


FIGURE 2. Assembly for controlling the outflow from the cell.

(*b*) It has been found to be unnecessary, with more slowly diffusing substances, to use the full capacity of the pipette (1·1 ml.); the slower rate of flow needed to form and maintain the boundary makes it possible to use as little as 0·4 ml. of solution for a determination.

(*c*) The slit has been used at a constant width of 0·0005 cm. and no attempt has been made to vary the intensity of the light reaching the focal plane by this means.

Thus, the contribution of the slit to the interference pattern recorded is constant. Kodak 0·250 plates have been used for exposure times corresponding to diffusion constants of 10^{-6} downwards. The photographic intensity has been controlled by introducing grey screens of known extinction in front of the source of light.

(*d*) The stop-assembly has been fitted with a vertical adjusting screw, for centering the stop on the boundary. Centering is done most conveniently by forming a boundary and then allowing spreading by diffusion to take place. As the boundary spreads beyond the confines of the stop, the innermost bands fade away; if the stop is not correctly centered, a blurring precedes their disappearance, due to the asymetrical distribution of light. The stop is therefore adjusted vertically, using visual inspection of the image, until the whole pattern appears quite sharp.

(2) *Materials*

A.R. glycine and KCl were used, solutions being made by weight of solid and volume of solution: sucrose was the ordinary 'granulated'. Lactoglobulin was prepared from whole fresh milk; first, by the method of Palmer (1934) and secondly by a combination of the methods of Palmer and of Sørensen (1939), described by Cecil & Ogston (1949).

Solutions of sucrose and KCl with lactoglobulin were made by dissolving weighed amounts in a solution of lactoglobulin already fully dialyzed against buffer. All measurements on lactoglobulin were made in buffer containing NaCl 0·1 M, Na acetate 0·1 M, acetic acid 0·04 M.

(3) *Analysis of homogeneity*

The values of X for a number of the inner minima may be measured; thence, values of $\log n_0/(n_j - \frac{3}{4})$ and of $\log \theta_{\text{max.}}/\theta_j$ $(= X_0/\beta X_j)$ may be calculated and compared with the curve given in figure 1. More precise information may be obtained, however, where $v > 15$, by measuring the values of X_j for those minima whose values of $\log n_0/(n_j - \frac{3}{4})$ bracket the values 0·2 and 0·9 and, by interpolation, obtaining the values of $\log \theta_{\text{max.}}/\theta_j$ corresponding to those values. Thence $\Delta_{0\cdot 2}$ and $\Delta_{0\cdot 9}$ may be obtained and, by interpolation in plots from tables 3 and 4, values of a^2 and α. Since the value of β used in the above calculation is the value for a single diffusing substance, the proper value of $\bar{\beta}$ may differ considerably from it. $\bar{\beta}/\beta_T$ can be obtained from the data of table 5, using the first-approximation values of a^2 and α; then the calculation is repeated, using the value of $\bar{\beta}$. A third approximation is rarely needed. From the final values of a^2 and α and the value of $\bar{D}$ obtained with $\bar{\beta}$, D_1 and D_2 are calculated from equation (6).

In measuring the inner minima $(\log n_0/(n_j - \frac{3}{4}) \approx 0\cdot 9)$ care is needed that measurement is made where the Fraunhofer fringes do not seriously interfere. In the case of a very rapidly diffusing impurity, care is also required that measurement is made at a time before that part of the boundary which gives rise to the bands measured has begun to spread beyond the limits of the stop; this is easily detected by fading of the intensity of the bands. In such a case it may be convenient to measure X_j at a time at which X_0 is not measurable; but the relationship of $1/X_0^2$ against t is accurately linear and extrapolation may be made with confidence.

RESULTS AND DISCUSSION

(1) *Test of the modified ray-theory for a single diffusing substance*

(*a*) *Distribution of bands.* Longsworth (1947) showed that the modified ray-theory gives the positions of all but the innermost bands correctly at rather long times after forming the boundary. This has been confirmed at much shorter times; figure 1 shows that the experimental points obtained from a range of concentrations of glycine, in relation to the theoretical curve. These points were obtained from measurements at various relative times and show that the relationship is independent of time; the only exception to this independence occurs where the apparent centre of a band is shifted by combination with the Fraunhofer system. In spite of exhaustive recrystallization and dialysis against buffer, it was not found possible to obtain a solution of lactoglobulin showing absolutely homogeneous behaviour (Cecil & Ogston 1949) though the points obtained usually lay close to the theoretical curve (figure 1).

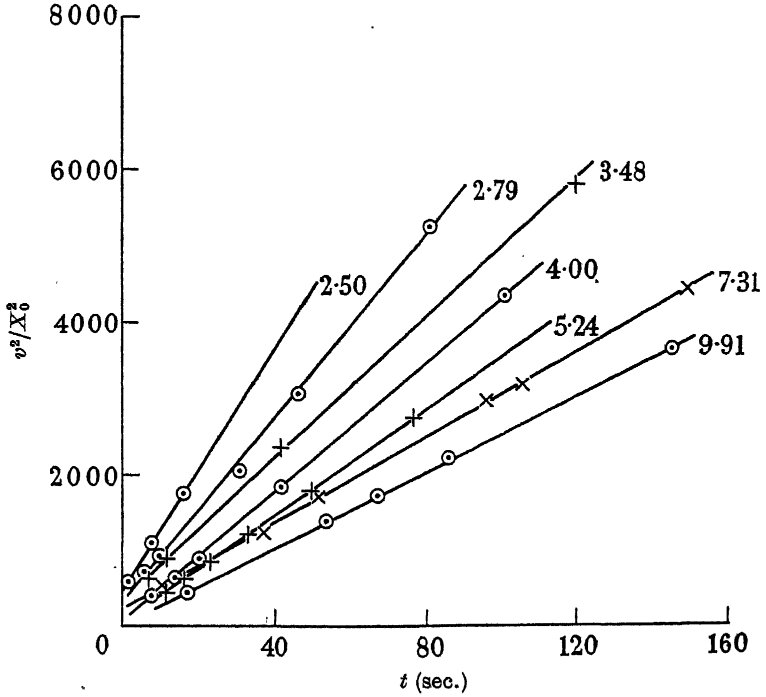


FIGURE 3. Linear plots of v^2/X_0^2 against t for solutions of glycine of varying v; values of v are shown against the lines; measurements were made at intersections between the boundary and Fraunhofer patterns.

(*b*) *Values of β.* A number of experiments were done with glycine at values of v down to 1·23, in the 1 cm. cell (i.e. down to 0·037 g./100 ml.). Down to $v = 2·5$, the method of measuring values of X_0 (the distance of the outermost dark band from the optic axis) and t corresponding to symmetrical intersections between the boundary and Fraunhofer patterns was used (see Theory 1 (*c*)). These values give fair linear plots of $1/X_0^2$ against t (figure 3). From their slopes and the values of β_T obtained by interpolation in table 2, the values of D_{20} given in table 6 were obtained.

These agree reasonably well with the values of D_{20} read from the curve given by Coulson *et al.* (1948).

At values of v lower than 2·5, the spacing of the Fraunhofer pattern is too wide to give a sufficient number of symmetrical intersections with the boundary pattern. An attempt to produce more intersections by using a wider stop (5 mm.) was unsuccessful, as the photographic definition proved to be insufficient. However, it was found that measurement of X_0 over certain ranges does lead to linear plots of $1/X_0^2$ against t (figure 4), though with too small a slope (a low slope was to be expected where a boundary band is 'running into' a Fraunhofer band). At lower values of v than 1·2, no linear plot could be obtained. From these observed slopes, the values of β required to give the known diffusion constants were obtained (table 7): the ratios β/β_T are plotted against v in figure 5 and fall on a smooth curve.

TABLE 6. VALUES OF D_{20} FOR GLYCINE

v	$D_{20}\times 10^7$ (obs.)	$D_{20}\times 10^7$ (from Coulson *et al.*)	difference
16·91	93·4	93·5	−0·1
11·33	93·6	94·2	−0·6
9·91	94·6	94·4	+0·2
8·55	93·9	94·6	−0·7
7·31	94·1	94·7	−0·6
5·61	94·4	95·0	−0·6
5·24	95·7	95·0	+0·7
4·00	97·1	95·2	+1·9
3·48	95·0	95·3	−0·3
2·79	94·9	95·5	−0·6
2·50	97·5	95·5	+2·0

TABLE 7. EMPIRICAL VALUES OF β FOR $v<2{\cdot}5$, FROM EXPERIMENTS ON GLYCINE

v	e	$D_{20}\times 10^7$ (from Coulson *et al.*)	β	β/β_T	range of θ_0 (radians × 10^4)
2·50	8·23	95·5	0·559	1·14	7–5
2·24	10·3	95·5	0·557	1·23	7–4
1·96	14·8	95·5	0·521	1·27	6–4
1·70	25·7	95·6	0·466	1·32	5–3
1·51	42·0	95·7	0·410	1·34	4–2
1·23	130	95·7	0·286	1·30	3–1

Thus it has proved possible to determine the value of D down to $v = 2{\cdot}5$, using the values of β_T given by the modified ray-theory, with an accuracy of 1 to 2 % and, further, to calibrate the method for measurement down to $v = 1{\cdot}2$. It is likely that the reproducibility of this calibration depends on the width of slit and size of stop used.

These sets of results together confirm the accuracy of β_T, obtained from the modified ray-theory, over a wide range. It is satisfactory that β_T begins to deviate seriously from β_B (table 2) at about the same value of v as was found empirically by Coulson *et al.* (1948).

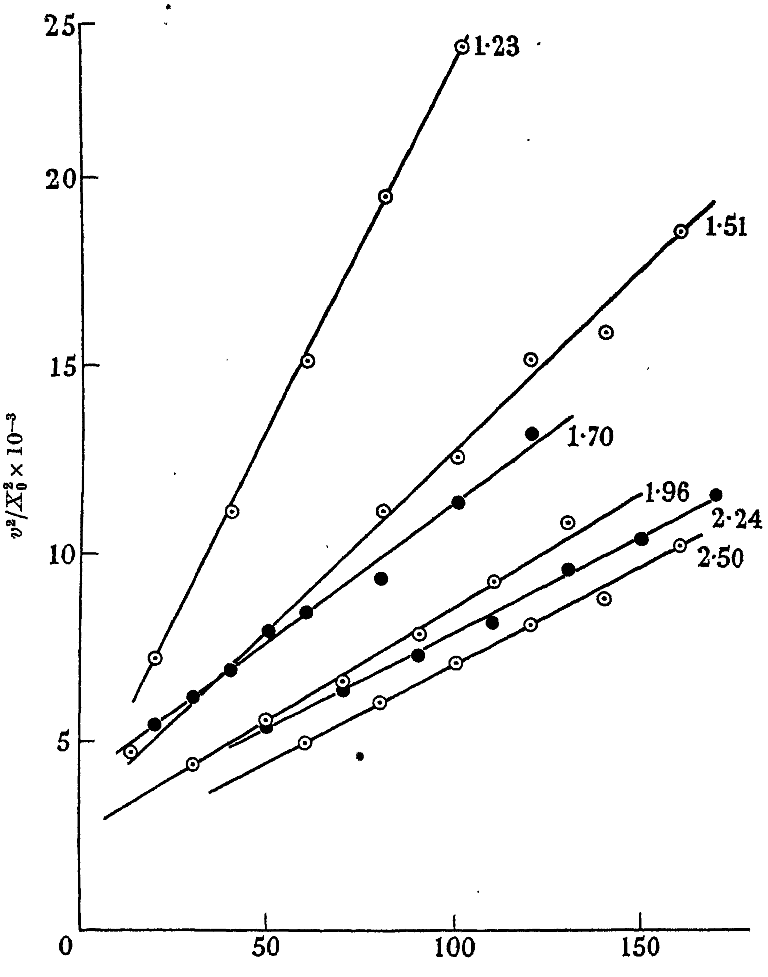


FIGURE 4. Approximately linear plots of v^2/X_0^2 against t for solutions of glycine at low values of v; values of v are shown against the lines. These measurements were not made at intersections between the boundary and Fraunhofer patterns and the slopes are anomalous.

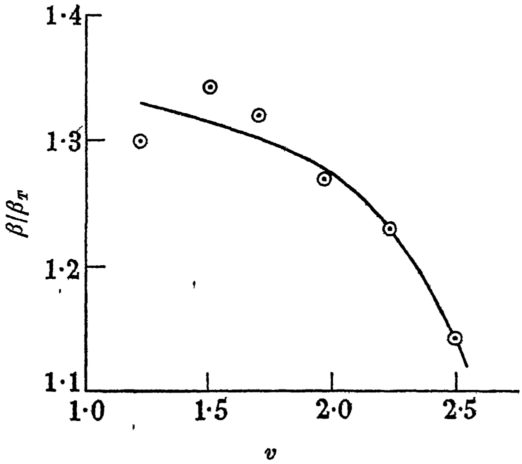


FIGURE 5. Deviation of β from theory at low values of v; data from the curves of figure 4.

(2) *Diffusion of heterogeneous material and of mixtures*

(*a*) *Heterogeneous samples of lactoglobulin.* The method described above has already been used to obtain the value (D_1) of the diffusion constant of a slightly contaminated main component in the case of lactoglobulin (Cecil & Ogston 1949). It has been applied also to some rather more heterogeneous samples of lactoglobulin which had been stored in solution for several months. The calculation of D_1 for these is set out in table 8; the values obtained are consistent with those obtained from the more homogeneous samples.

TABLE 8. CORRECTION OF THE APPARENT DIFFUSION CONSTANTS OF HETEROGENEOUS SAMPLES OF LACTOGLOBULIN

v (total)	$\Delta_{0\cdot2}$	$\Delta_{0\cdot9}$	α	a^2	$\bar{D}\times10^7$	$D_1\times10^7$
37·39	0·004	0·032	0·98	0·11	7·59	7·38
37·19	0·004	0·038	0·99	0·03	7·56	7·44
36·93	0·004	0·035	0·98	0·11	7·64	7·45
15·97	0·007	0·055	0·96	0·11	7·67	7·27
15·12	0·008	0·076	0·97	0·04	7·90	7·52

mean D_1 for $v = 33$ (Cecil & Ogston 1949) = 7·34

(*b*) *Mixtures of two diffusing components.* In order to test the method more strictly, the diffusion of mixtures of lactoglobulin with sucrose and with potassium chloride was observed. The analysis is set out in table 9. In both cases, the agreement of the quantities estimated from the analysis with their known values is satisfactory.

TABLE 9. ANALYSIS OF THE DIFFUSION OF KNOWN MIXTURES

	1st approximation	2nd approximation	known values
	(*a*) *Lactoglobulin and sucrose*, $v_1+v_2 = 33\cdot26$		
$\bar{D}\times10^7$	8·43	8·40	—
$\Delta_{0\cdot2}$	0·0115	0·0121	—
$\Delta_{0\cdot9}$	0·080	0·081	—
α	0·92	0·905	0·904
a^2	0·17	0·18	0·167
$\bar{\beta}/\beta_T$	0·998$_5$	0·998	—
$D_1\times10^7$	—	7·53	7·34
$D_2\times10^7$	—	42	44
	(*b*) *Lactoglobulin and potassium chloride*, $v_1+v_2 = 40\cdot89$		
$\bar{D}\times10^7$	10·61	10·44	—
$\Delta_{0\cdot2}$	0·046	0·0495	—
$\Delta_{0\cdot9}$	0·403	0·406	—
α	0·83	0·81	0·81
a^2	0·048	0·050	0·049
$\bar{\beta}/\beta_T$	0·992	0·992	—
$D_1\times10^7$	—	7·58	7·34
$D_2\times10^7$	—	152	c. 150

CONCLUSIONS

The application of the modified ray-theory to the analysis of the interference pattern arising from a diffusion boundary is justified by the results given above. By its use, an accurate analysis of the diffusion pattern of a mixture of two diffusing components in terms of their proportions and diffusion constants can be performed; also correction can be made with fair accuracy for the effect of heterogeneous material on the apparent diffusion constant of a main component. It should be pointed out that the method of analysis described does not make nearly full use of the experimental data provided by the record of a diffusion experiment and that its scope and accuracy might be increased considerably if a less clumsy means of handling the equations were to be found.

REFERENCES

Cecil, R. & Ogston, A. G. 1949 *Biochem. J.* **43**, 592.
Coulson, C. A., Cox, J. T., Ogston, A. G. & Philpot, J. St L. 1948 *Proc. Roy. Soc.* A, **192**, 382.
Kegeles, G. & Gosting, L. J. 1947 *J. Amer. Chem. Soc.* **69**, 2516.
Longsworth, L. G. 1947 *J. Amer. Chem. Soc.* **69**, 2510.
Palmer, A. H. 1934 *J. Biol. Chem.* **104**, 359.
Sørensen, S. P. L. 1939 *C.R. Lab. Carlsberg*, **23**, 55.

A systematic treatment of moving axes in hydrodynamics

BY G. C. MCVITTIE, *Queen Mary College, London*

(*Communicated by G. Temple, F.R.S.—Received* 3 *August* 1948)

A method is developed for transforming the equations of hydrodynamics to a system of curvilinear co-ordinates in motion relative to fixed axes using the tensor-calculus and without employing Coriolis's theorem. The basic entity is the kinetic metric, a four-dimensional quadratic form defined through the kinetic energy of unit mass of fluid. The mechanics of special relativity are used to obtain, by approximation in terms of $1/c^2$, where c is the velocity of light, the classical formulae. The equations of motion in terms of the energy-tensor, the four-dimensional vorticity-tensor and the velocity-components are successively obtained and the equation of continuity is shown to be independent (in mathematical form) of the motion of the co-ordinate-system. This property holds also for the equation of heat-transfer in a non-viscous fluid. Applications are made to the case of local Cartesian and local cylindrical polar co-ordinates on the Earth's surface. Formulae for the rate of change of vorticity due to Helmholtz and to Petterssen, respectively, are obtained as special cases and Sawyer's theory of tropical cyclones is also discussed.

1. INTRODUCTION

In classical hydrodynamics, the use of curvilinear spatial co-ordinates is well-known. The conversion of the equations to these co-ordinates depends ultimately on (i) the existence of a unique 'time' in Newtonian mechanics common to all observers; and (ii) on the possibility of introducing spatial co-ordinate-transformations which do not involve the time explicitly. If we denote by (X^1, X^2, X^3) the fixed rectangular co-ordinates* of an element of fluid and by X^4 the absolute time, then the transformation to a curvilinear system (x^1, x^2, x^3) is given by

$$\left.\begin{aligned} X^r &= f^r(x^1, x^2, x^3) \quad (r = 1, 2, 3), \\ X^4 &= x^4. \end{aligned}\right\} \qquad (1\cdot1)$$

---

* By this we mean that the equations of motion of a particle of mass m acted on by a force (P^1, P^2, P^3) have the form $m d^2X^r/(dX^4)^2 = P^r$, $(r = 1, 2, 3)$.

The tensor calculus in three dimensions will then convert the equations of hydrodynamics into the requisite form. But these operations fail as soon as we consider curvilinear, or indeed rectangular, co-ordinate systems which are in motion relative to the fixed rectangular co-ordinates. Instead of (1·1) we now have transformation-equations of the form

$$\left.\begin{aligned} X^r &= h^r(x^1, x^2, x^3, x^4) \quad (r = 1, 2, 3), \\ X^4 &= x^4, \end{aligned}\right\} \tag{1·2}$$

and, to obtain the equations of hydrodynamics, we have to take refuge in cumbersome and unsymmetrical operations based on Coriolis's theorem. In this paper we develop a method of establishing the equations of hydrodynamics for co-ordinate-transformations of type (1·2) which is parallel to the classical theory developed for those of type (1·1).

The requisite mathematical technique is indeed ready to our hand in the extensions of Newtonian mechanics found in the theory of special relativity. It is true that attention is there concentrated on velocities which are appreciable fractions of that of light, so that the time-co-ordinates cannot be related by simple identity as in (1·2). But this qualification is not essential to the mathematical formulation. Moreover the degeneration of special relativity into Newtonian mechanics is easily performed by successive approximations in terms of $1/c^2$, where c is the velocity of light.

2. THE KINETIC METRIC

Consider the kinetic energy of unit mass of fluid which, relative to the fixed rectangular axes, is

$$K = \frac{1}{2}\left\{\left(\frac{dX^1}{dX^4}\right)^2 + \left(\frac{dX^2}{dX^4}\right)^2 + \left(\frac{dX^3}{dX^4}\right)^2\right\}. \tag{2·1}$$

We can imagine a geometrical representation of the motion of the unit mass, its position (X^1, X^2, X^3) at time X^4 being plotted as a 'point' (X^1, X^2, X^3, X^4) in a four-dimensional manifold whose metric is

$$ds^2 = (dX^4)^2 - \frac{1}{c^2}\{(dX^1)^2 + (dX^2)^2 + (dX^3)^2\}. \tag{2·2}$$

The history of the motion of the unit mass will be represented by a curve in the four-dimensional manifold along which

$$ds^2 = \{1 - 2K/c^2\}(dX^4)^2. \tag{2·3}$$

We do not suggest that the manifold has any physical significance: the representation is merely an aid to calculation. We call ds^2, the *kinetic metric*, and its transformations under (1·2) will form the basis of our method. From (2·3) we observe that, neglecting terms in $1/c^2$, we have $ds \equiv dX^4$ for the motion of the unit mass and hence, in Newtonian Mechanics, s and X^4 may be regarded as identical.

It is convenient to introduce the convention that an index represented by a Greek letter can take the values 1 to 4 whilst one represented by a Roman letter can take the values 1 to 3 only. Using the summation convention, the differential form of (1·2) is

$$\left.\begin{aligned} dX^r &= \frac{\partial h^r}{\partial x^\mu}dx^\mu, \\ dX^4 &= dx^4. \end{aligned}\right\} \tag{2·4}$$

Hence
$$K = \frac{1}{2}\left\{\gamma_{44}+2\gamma_{p4}\frac{dx^p}{dx^4}+\gamma_{pq}\frac{dx^p}{dx^4}\frac{dx^q}{dx^4}\right\}, \tag{2·5}$$

where
$$\left.\begin{aligned} \gamma_{44} &= \sum_{r=1}^{3}\left(\frac{\partial h^r}{\partial x^4}\right)^2, \quad \gamma_{p4} = \gamma_{4p} = \sum_{r=1}^{3}\frac{\partial h^r}{\partial x^p}\frac{\partial h^r}{\partial x^4}, \\ \gamma_{pq} &= \gamma_{qp} = \sum_{r=1}^{3}\frac{\partial h^r}{\partial x^p}\frac{\partial h^r}{\partial x^q}. \end{aligned}\right\} \tag{2·6}$$

Hence (2·2) becomes

$$ds^2 = \left(1-\frac{\gamma_{44}}{c^2}\right)(dx^4)^2-\frac{1}{c^2}\{\gamma_{pq}dx^p\,dx^q+2\gamma_{p4}dx^p\,dx^4\}, \tag{2·7}$$

and, if we write,
$$\left.\begin{aligned} g_{44} &= 1-\frac{\gamma_{44}}{c^2}, \quad g_{p4} = g_{4p} = -\frac{\gamma_{p4}}{c^2}, \\ g_{pq} &= g_{qp} = -\frac{\gamma_{pq}}{c^2}, \end{aligned}\right\} \tag{2·8}$$

(2·7) becomes
$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu. \tag{2·9}$$

Since ds^2 is an invariant, $g_{\mu\nu}$ forms a symmetric covariant tensor of rank 2 in four dimensions. Clearly, by (2·6) and (2·8), the $\gamma_{\mu\nu}$ and $g_{\mu\nu}$ are, in general, functions of the time x^4 as well as of the spatial co-ordinates. Also if the curvilinear co-ordinate system (x^1, x^2, x^3) is 'at rest' relative to the fixed rectangular system, the functions h^r in (1·2) do not involve x^4 explicitly. Hence, in this case,

$$\left.\begin{aligned} \gamma_{44} &= 0, \quad \gamma_{p4} = \gamma_{4p} = 0, \\ g_{44} &= 1, \quad g_{p4} = g_{4p} = 0, \end{aligned}\right\} \tag{2·10}$$

which are the conditions that the theory based on (2·9) should reduce to the classical theory derived from equations (1·1).

In Newtonian hydrodynamics we deal with velocities which are small compared with c. We therefore assume that the functions h^r are such that the $\gamma_{\mu\nu}$ do not contain a factor of order c^2. Hence in (2·8), g_{44} is of order unity whilst the remaining $g_{\mu\nu}$ are of order $1/c^2$.

We shall also require the expression for the determinant of the tensor $g_{\mu\nu}$. We write $g \equiv \|g_{\mu\nu}\|$ and $\Delta_0 \equiv \|\gamma_{\mu\nu}\|$. If Δ denotes the cofactor of γ_{44}, and Δ_{4p}, the cofactor of γ_{4p} in Δ_0, then on evaluating g we find

$$g = -\frac{\Delta}{c^6}+\frac{\Delta_0}{c^8}.$$

Hence, to terms of order $1/c^5$, we find

$$\sqrt{(-g)} = \frac{\sqrt{\Delta}}{c^3}\left(1-\frac{1}{2}\frac{1}{c^2}\frac{\Delta_0}{\Delta}\right). \tag{2·11}$$

We can now express the contravariant form of $g_{\mu\nu}$ in terms of the contravariant $\gamma_{\mu\nu}$. We define the contravariant forms by

$$g^{\mu\nu} = \frac{\text{Cofactor of } g_{\mu\nu} \text{ in } g}{g},$$

and

$$\gamma^{pq} = \frac{\text{Cofactor of } \gamma_{pq} \text{ in } \Delta}{\Delta},$$

and we find

$$\left.\begin{aligned} g^{44} &= \{1+O(1/c^2)\}, \\ g^{p4} = g^{4p} &= \Delta_{4p}\{1+O(1/c^2)\}/\Delta, \\ g^{pq} &= -c^2\{\gamma^{pq}+O(1/c^2)\}. \end{aligned}\right\} \tag{2·12}$$

We observe that g^{44}, g^{p4} are of order unity whilst the g^{pq} are of order c^2. Lastly, if the co-ordinate system we are employing is itself the fixed rectangular system we have, comparing (2·9) with (2·2) and using (2·12) also,

$$\left.\begin{aligned} &g_{44} = 1, \quad g_{pp} = -\frac{1}{c^2}, \quad \sqrt{(-g)} = \frac{1}{c^3}, \\ &g^{44} = 1, \quad g^{pp} = -c^2, \end{aligned}\right\} \tag{2·13}$$

all the other $g_{\mu\nu}$ and $g^{\mu\nu}$ being zero.

3. DIVERGENCE AND VORTICITY OF THE FLUID

We define the contravariant velocity four-vector of the fluid element by

$$v^{\mu} = \frac{dx^{\mu}}{ds}. \tag{3·1}$$

Since in Newtonian mechanics we neglect terms in $1/c^2$ and $ds = dx^4$, (3·1) reduces to a four-vector u^{μ} whose components are

$$u^4 = 1, \quad u^p = \frac{dx^p}{dx^4}. \tag{3·2}$$

For the special case of fixed rectangular co-ordinates, we write this four-vector as

$$U^4 = 1, \quad U^p = \frac{dX^p}{dX^4}, \tag{3·3}$$

and (U^1, U^2, U^3) are the velocity-components usually denoted by (u, v, w) in the text-books.

The four-dimensional divergence (an invariant) of the vector v^{μ} is

$$\Theta = \frac{1}{\sqrt{(-g)}} \frac{\partial(\sqrt{(-g)}\, v^{\mu})}{\partial x^{\mu}},$$

which, on using (2·11) and (3·2) and neglecting terms of order $1/c^2$, gives us in Newtonian mechanics

$$\Theta = \frac{1}{\sqrt{\Delta}}\left\{\frac{\partial\sqrt{\Delta}}{\partial x^4} + \frac{\partial(\sqrt{\Delta}\, u^p)}{\partial x^p}\right\}. \tag{3·4}$$

For fixed rectangular axes, by (3·3) and (2·13), we have

$$\Theta = \frac{\partial U^p}{\partial X^p},$$

which is the well-known expression for the quantity variously known in hydrodynamics as the 'expansion', the 'dilatation' or the 'divergence' of the fluid. We shall adopt the latter term which is current in dynamical meteorology. In most practical cases, the determinant Δ does not involve the time x^4 explicitly and (3·4) reduces to

$$\Theta = \frac{\partial u^p}{\partial x^p} + \tfrac{1}{2} u^p \frac{\partial \log \Delta}{\partial x^p}. \tag{3·5}$$

We note that, since Θ does not involve γ_{44} or the γ_{p4}, its mathematical form is now independent of the motion of the co-ordinate-system.

Turning to the definition of vorticity we remember that, in the three-dimensional treatment of hydrodynamics based on (1·1), vorticity appears as a solenoidal three-vector (McConnell 1931). Such vectors, when we pass to a four-dimensional treatment, become portions of antisymmetric tensors of rank 2. Moreover, vorticity is a property of the motion of a fluid whose *quantity*, as opposed to its *intensity*, is relevant, which leads us to a tensor-density as the fundamental definition (Eddington 1924). We write

$$v_\mu = g_{\mu\nu} v^\nu, \tag{3·6}$$

and we denote by $v_{\mu,\nu}$ the covariant derivative of v_μ (McConnell 1931) with respect to the kinetic metric (2·9). The six quantities $(v_{\mu,\nu} - v_{\nu,\mu})$ then form the components of an antisymmetric tensor of rank 2 and we shall define the *vorticity-tensor* $\xi_{\mu\nu}$ by means of the relation

$$c\sqrt{(-g)}\,\xi_{\mu\nu} = v_{\mu,\nu} - v_{\nu,\mu} \equiv \frac{\partial v_\mu}{\partial x^\nu} - \frac{\partial v_\nu}{\partial x^\mu}. \tag{3·7}$$

Before discussing the general case, let us suppose that our co-ordinate system is the fixed rectangular system in which (2·13) and (3·3) hold good. We find, from (3·7),

$$(\xi_{41}, \xi_{42}, \xi_{43}) = \left(\frac{\partial U^1}{\partial X^4}, \frac{\partial U^2}{\partial X^4}, \frac{\partial U^3}{\partial X^4}\right),$$

$$(\xi_{23}, \xi_{31}, \xi_{12}) = \left\{\left(\frac{\partial U^3}{\partial X^2} - \frac{\partial U^2}{\partial X^3}\right), \left(\frac{\partial U^1}{\partial X^3} - \frac{\partial U^3}{\partial X^1}\right), \left(\frac{\partial U^2}{\partial X^1} - \frac{\partial U^1}{\partial X^2}\right)\right\},$$

and we observe that the three 'spatial' components of $\xi_{\mu\nu}$ are the components of (absolute) vorticity (ξ, η, ζ) which thus find themselves associated in a single tensor with the partial derivatives of the velocity-components with respect to the time.

We return to the reduction to Newtonian mechanics of the general form (3·7). An alternative way of writing (2·8) using the Kroenecker delta, $\delta_{\mu\nu}$, is

$$g_{\mu\nu} = \delta_{\mu4}\delta_{\nu4} - \frac{1}{c^2}\gamma_{\mu\nu}.$$

Hence (3·7) and (3·6) give

$$c\sqrt{(-g)}\,\xi_{\mu\nu} = \frac{\partial}{\partial x^\nu}\left\{\left(\delta_{\mu4}\delta_{\sigma4}-\frac{1}{c^2}\gamma_{\mu\sigma}\right)v^\sigma\right\} - \frac{\partial}{\partial x^\mu}\left\{\left(\delta_{\nu4}\delta_{\sigma4}-\frac{1}{c^2}\gamma_{\nu\sigma}\right)v^\sigma\right\}$$

$$= \delta_{\mu4}\frac{\partial v^4}{\partial x^\nu} - \delta_{\nu4}\frac{\partial v^4}{\partial x^\mu} - \frac{1}{c^2}\left\{\frac{\partial(\gamma_{\mu\sigma}v^\sigma)}{\partial x^\nu} - \frac{\partial(\gamma_{\nu\sigma}v^\sigma)}{\partial x^\mu}\right\}. \tag{3·8}$$

The transition to Newtonian mechanics is now made via (2·11) and (3·2) by equating terms of order $1/c^2$ on each side of (3·8), neglecting terms of higher order. We find

$$\sqrt{\Delta}\,\xi_{\mu\nu} = -\left(\frac{\partial\gamma_{\mu4}}{\partial x^\nu} - \frac{\partial\gamma_{\nu4}}{\partial x^\mu}\right) - \left\{\frac{\partial(\gamma_{\mu p}u^p)}{\partial x^\nu} - \frac{\partial(\gamma_{\nu p}u^p)}{\partial x^\mu}\right\}, \tag{3·9}$$

and we therefore write
$$\xi_{\mu\nu} = \omega_{\mu\nu} + \Omega_{\mu\nu}, \tag{3·10}$$

where
$$\sqrt{\Delta}\,\omega_{\mu\nu} = -\left(\frac{\partial\gamma_{\mu4}}{\partial x^\nu} - \frac{\partial\gamma_{\nu4}}{\partial x^\mu}\right), \quad \sqrt{\Delta}\,\Omega_{\mu\nu} = -\left\{\frac{\partial(\gamma_{\mu p}u^p)}{\partial x^\nu} - \frac{\partial(\gamma_{\nu p}u^p)}{\partial x^\mu}\right\}. \tag{3·11}$$

The three components of absolute vorticity as ordinarily understood in hydrodynamics are, as we have seen, ξ_{23}, ξ_{31} and ξ_{12}. We can simplify the notation for these three components as follows: let *lmn* be any cyclic permutation of the indices 123 and let

$$\xi_l = \xi_{mn}, \quad \omega_l = \omega_{mn}, \quad \Omega_l = \Omega_{mn}. \tag{3·12}$$

Then
$$\left.\begin{aligned} \xi_l &= \omega_l + \Omega_l, \\ \text{where}\quad \sqrt{\Delta}\,\omega_l &= -\left(\frac{\partial\gamma_{m4}}{\partial x^n} - \frac{\partial\gamma_{n4}}{\partial x^m}\right), \\ \sqrt{\Delta}\,\Omega_l &= -\left\{\frac{\partial(\gamma_{mp}u^p)}{\partial x^n} - \frac{\partial(\gamma_{np}u^p)}{\partial x^m}\right\}. \end{aligned}\right\} \tag{3·13}$$

We observe that the ω_l $(l = 1, 2, 3)$ do not involve the velocity-components of the fluid; they represent the contribution to the absolute vorticity due to the motion of the co-ordinate-system itself. We may call Ω_l $(l = 1, 2, 3)$ the *relative vorticity*, or simply the *vorticity*, of the fluid, since it is the vorticity as it would appear to an observer using the co-ordinate system (x^1, x^2, x^3). If the motion of the fluid appears to be irrotational to such an observer, i.e. if

$$\gamma_{qp}u^p = \frac{\partial\phi(x^1, x^2, x^3)}{\partial x^q} \quad (q = 1, 2, 3),$$

then $\Omega_l = 0$ $(l = 1, 2, 3)$. But, of course, this is not sufficient to make the absolute vorticity vanish when the co-ordinate-system is itself in motion.

4. EQUATIONS OF MOTION AND OF CONTINUITY

In this section we recapitulate briefly the generalizations of the equations of motion and of continuity of a fluid in special relativity. We begin with classical hydrodynamics referred to the fixed rectangular co-ordinates. If ρ is the density and p the pressure of the fluid, and if the external force per unit volume other than the pressure-gradient is (X, Y, Z), we can write the equations of motion in the form

$$\left.\begin{aligned}\frac{\partial\rho U^1}{\partial X^4}+\frac{\partial}{\partial X^1}\{\rho(U^1)^2+p\}+\frac{\partial\rho U^1U^2}{\partial X^2}+\frac{\partial\rho U^1U^3}{\partial X^3}&=\rho X,\\ \frac{\partial\rho U^2}{\partial X^4}+\frac{\partial\rho U^1U^2}{\partial X^1}+\frac{\partial}{\partial X^2}\{\rho(U^2)^2+p\}+\frac{\partial\rho U^2U^3}{\partial X^3}&=\rho Y,\\ \frac{\partial\rho U^3}{\partial X^4}+\frac{\partial\rho U^1U^3}{\partial X^1}+\frac{\partial\rho U^2U^3}{\partial X^2}+\frac{\partial}{\partial X^3}\{\rho(U^3)^2+p\}&=\rho Z,\end{aligned}\right\} \tag{4·1}$$

whilst the equation of continuity is

$$\frac{\partial\rho}{\partial X^4}+\frac{\partial\rho U^1}{\partial X^1}+\frac{\partial\rho U^2}{\partial X^2}+\frac{\partial\rho U^3}{\partial X^3}=0. \tag{4·2}$$

Remembering that, by (3·3), $U^4 = 1$ and using (2·13), we can define a four-dimensional energy-tensor

$$T^{\mu\nu}=\rho U^\mu U^\nu-\frac{g^{\mu\nu}}{c^2}p, \tag{4·3}$$

whose components are the quantities which occur in the partial derivatives in the left-hand sides of (4·1) and (4·2). The only proviso we must make is that, to obtain the term $\partial\rho/\partial X^4$ in (4·2) from $\partial T^{44}/\partial X^4$ we must, in the latter, neglect the term in $-p/c^2$ in accordance with our Newtonian approximation. The four equations (4·1) and (4·2) may then be written

$$\frac{\partial T^{\mu\nu}}{\partial X^\nu}=(\rho X,\rho Y,\rho Z,0), \tag{4·4}$$

according as $\mu = 1, 2, 3$, or 4. The generalization of (4·3) and (4·4) to any co-ordinate system (x^1, x^2, x^3, x^4) in terms of which the kinetic metric has the form (2·9) is immediate, since they are tensor equations. We regard ρ and p as invariants and we replace (4·3) by

$$T^{\mu\nu}=\rho v^\mu v^\nu-\frac{g^{\mu\nu}}{c^2}p, \tag{4·5}$$

whilst the left-hand sides of the four equations (4·4) become

$$T^{\mu\nu}_{,\nu}=\frac{1}{\sqrt{(-g)}}\frac{\partial\{\sqrt{(-g)}\,T^{\mu\nu}\}}{\partial x^\nu}+\left\{{\mu \atop \sigma\nu}\right\}T^{\sigma\nu}, \tag{4·6}$$

i.e. they reduce to the four components of a contravariant four-vector $T^{\mu\nu}_{,\nu}$ called the 'vectorial divergence' of the energy-tensor. We must generalize the Newtonian force per unit volume (X, Y, Z) to a four-vector f^μ so that instead of (4·4) we now have

$$T^{\mu\nu}_{,\nu}=\rho f^\mu \tag{4·7}$$

as the equations of motion and of continuity of the fluid. We can, however, prove as follows that f^4 is always zero: If the co-ordinates are the fixed rectangular, we denote the force four-vector by f_0^4 and we know that $f_0^4 = 0$ by (4·4). But by the rule for transformation of contravariant vectors we have by (1·2),

$$f^4=\frac{\partial x^4}{\partial X^\nu}f_0^\nu=\frac{\partial x^4}{\partial X^4}f_0^4=f_0^4=0.$$

5. EQUATIONS OF MOTION IN TERMS OF VORTICITY-TENSOR

In covariant differentiation we treat $g_{\mu\nu}$ and $g^{\mu\nu}$ as if they were constants and we know that the covariant derivative, $I_{,\nu}$, of an invariant I is simply $\partial I/\partial x^\nu$. Hence applying the operation of covariant differentiation to the expression (4·5) for $T^{\mu\nu}$ and inserting the result into (4·7), we find

$$v^\mu(\rho v^\nu)_{,\nu}+\rho v^\nu v^\mu_{,\nu}-\frac{g^{\mu\nu}}{c^2}\frac{\partial p}{\partial x^\nu}=\rho f^\mu. \tag{5·1}$$

If we note that $v^\mu_{,\nu}=g^{\mu\sigma}v_{\sigma,\nu}$ and use (3·7) we can write (5·1) as

$$v^\mu(\rho v^\nu)_{,\nu}+c\sqrt{(-g)}\,\xi_{\sigma\nu}g^{\mu\sigma}\rho v^\nu+g^{\mu\sigma}\rho v^\nu v_{\nu,\sigma}-\frac{g^{\mu\sigma}}{c^2}\frac{\partial p}{\partial x^\sigma}=\rho f^\mu. \tag{5·2}$$

We introduce the invariant $$(v)^2=v^\nu v_\nu, \tag{5·3}$$

and it is easy to show that $$\frac{\partial(v)^2}{\partial x^\sigma}=2v^\nu v_{\nu,\sigma}. \tag{5·4}$$

Hence (5·2) becomes

$$v^\mu(\rho v^\nu)_{,\nu}+g^{\mu\sigma}\left\{c\sqrt{(-g)}\,\xi_{\sigma\nu}\rho v^\nu+\tfrac{1}{2}\rho\frac{\partial(v)^2}{\partial x^\sigma}-\frac{1}{c^2}\frac{\partial p}{\partial x^\sigma}\right\}=\rho f^\mu.$$

We multiply this equation by $g_{\mu\tau}$ and sum over the four values of μ; we find finally

$$v_\tau(\rho v^\nu)_{,\nu}+c\sqrt{(-g)}\,\xi_{\tau\nu}\rho v^\nu+\tfrac{1}{2}\rho\frac{\partial(v)^2}{\partial x^\tau}-\frac{1}{c^2}\frac{\partial p}{\partial x^\tau}=\rho f_\tau. \tag{5·5}$$

We now proceed to the Newtonian approximation. By (3·2) and (2·8) we have the following approximate formulae:

$$\left.\begin{aligned}v_4&=1-\frac{\gamma_{44}+\gamma_{4q}u^q}{c^2},\quad v_q=-\frac{\gamma_{q4}+\gamma_{qr}u^r}{c^2},\\(v)^2&=g_{\nu\sigma}v^\nu v^\sigma=1-\frac{1}{c^2}\{\gamma_{44}+2\gamma_{q4}u^q+(u)^2\},\end{aligned}\right\} \tag{5·6}$$

where $$(u)^2=\gamma_{pq}u^pu^q. \tag{5·7}$$

Again, $c\sqrt{(-g)}\,\xi_{\mu\nu}=\sqrt{\Delta}\,\xi_{\mu\nu}$, the value of which is now given by (3·9); and

$$(\rho v^\nu)_{,\nu}=(\rho u^\nu)_{,\nu}=\frac{1}{\sqrt{\Delta}}\left\{\frac{\partial(\rho\sqrt{\Delta})}{\partial x^4}+\frac{\partial(\rho\sqrt{\Delta}\,u^q)}{\partial x^q}\right\}. \tag{5·8}$$

Lastly, by definition, $f_\tau=g_{\tau\sigma}f^\sigma=g_{\tau r}f^r$, since $f^4=0$.

Hence by (2·8), we write

$$f_4=-\frac{1}{c^2}\gamma_{4r}f^r=-\frac{1}{c^2}F_4,\quad f_q=-\frac{1}{c^2}\gamma_{qr}f^r=-\frac{1}{c^2}F_q, \tag{5·9}$$

where F_q $(q=1,2,3)$ are the components of the Newtonian force per unit volume with respect to the, in general curvilinear, co-ordinate-system (x^1,x^2,x^3). We introduce the foregoing expressions into (5·5) and we find when $\tau=4$:

$$(\rho u^\nu)_{,\nu}\left\{1-O\left(\frac{1}{c^2}\right)\right\}+\frac{1}{c^2}\left[\sqrt{\Delta}\,\rho u^p\xi_{4p}-\tfrac{1}{2}\rho\frac{\partial}{\partial x^4}\{\gamma_{44}+2\gamma_{p4}u^p+(u)^2\}-\frac{\partial p}{\partial x^4}\right]=-\rho\frac{F_4}{c^2}; \tag{5·10}$$

when $\tau = q = 1$, 2 or 3, eliminating the factor $1/c^2$:

$$-(\gamma_{q4}+\gamma_{qr}u^r)(\rho u^\nu)_{,\nu}+\sqrt{\Delta}\,\rho(\xi_{q4}+u^p\xi_{qp})-\tfrac{1}{2}\rho\frac{\partial}{\partial x^q}\{\gamma_{44}+2\gamma_{p4}u^p+(u)^2\}-\frac{\partial p}{\partial x^q}=-\rho F_q, \tag{5·11}$$

where in the last two equations the $\xi_{\mu\nu}$ are defined by (3·9). Hence if in (5·10) we neglect the terms in $1/c^2$, we obtain

$$(\rho u^\nu)_{,\nu}=0, \tag{5·12}$$

which by (5·8) gives us the equation of continuity

$$\frac{\partial(\rho\sqrt{\Delta})}{\partial x^4}+\frac{\partial(\rho\sqrt{\Delta}\,u^p)}{\partial x^p}=0.$$

Incidentally we observe that γ_{44} and the γ_{p4} are not involved in this equation so that its mathematical form is the same whether the co-ordinate-system is in motion or not; indeed the motion can only affect the equation of continuity in cases where Δ is an explicit function of the time.

In the three equations (5·11), since the first term on the left now vanishes, we must have, on reversing the order of the indices in the $\xi_{\mu\nu}$,

$$\sqrt{\Delta}\,(\xi_{4q}+u^p\xi_{pq})+\frac{1}{2}\frac{\partial}{\partial x^q}\{\gamma_{44}+2\gamma_{p4}u^p+(u)^2\}=-\frac{1}{\rho}\frac{\partial p}{\partial x^q}+F_q\quad(q=1,2,3), \tag{5·13}$$

which are the required equations of motion expressed in terms of the vorticity-tensor (3·9). If the co-ordinate-system is at rest relative to the fixed rectangular axes, the terms in γ_{44}, γ_{p4} are zero. The equations (5·13) may also, of course, be expressed in terms of the velocity-components u^p alone. The case of most interest occurs for spatially orthogonal co-ordinate-systems in which $\gamma_{pq}=0$ when $p\neq q$. Using (3·9) and (5·7) and writing

$$\frac{d}{dx^4}=\frac{\partial}{\partial x^4}+u^p\frac{\partial}{\partial x^p}, \tag{5·14}$$

we obtain, suspending the summation convention,

$$\begin{aligned}\frac{d(\gamma_{qq}u^q)}{dx^4}&-\frac{1}{2}\sum_{p=1}^{3}\frac{\partial\gamma_{pp}}{\partial x^q}(u^p)^2\\&-\left(\frac{1}{2}\frac{\partial\gamma_{44}}{\partial x^q}-\frac{\partial\gamma_{q4}}{\partial x^4}\right)-\sum_{p=1}^{3}\left(\frac{\partial\gamma_{p4}}{\partial x^q}-\frac{\partial\gamma_{q4}}{\partial x^p}\right)u^p=-\frac{1}{\rho}\frac{\partial p}{\partial x^q}+F_q.\end{aligned} \tag{5·15}$$

6. RATE OF CHANGE OF THE VORTICITY-TENSOR

We return to the equations (5·5) which we now write in the form

$$\begin{aligned}c\sqrt{(-g)}\,v^\nu\xi_{\tau\nu}+\frac{1}{2}\frac{\partial(v)^2}{\partial x^\tau}&=-\frac{v_\tau}{\rho}(\rho v^\nu)_{,\nu}+\frac{1}{c^2}\frac{1}{\rho}\frac{\partial p}{\partial x^\tau}+f_\tau\\&=g_\tau\quad\text{(say)}.\end{aligned} \tag{6·1}$$

The four functions g_τ form the components of a covariant four-vector since v_τ, $\partial p/\partial x^\tau$ and f_τ are covariant four-vectors, whilst ρ and $(\rho v^\nu)_{,\nu}$ are invariants. By applying covariant differentiation to g_τ, we can construct from its covariant derivatives an antisymmetric covariant tensor of rank 2, viz:

$$g_{\tau,\mu}-g_{\mu,\tau}=\frac{\partial g_\tau}{\partial x^\mu}-\frac{\partial g_\mu}{\partial x^\tau}. \tag{6·2}$$

In the left-hand side of (6·1), $\sqrt{(-g)}$ is a function of the $g_{\mu\nu}$ and so may be treated as a constant in covariant differentiation. Hence we obtain, using also (3·7),

$$g_{\tau,\mu}-g_{\mu,\tau}=c\sqrt{(-g)}\left\{\frac{\partial v^\nu}{\partial x^\mu}\xi_{\tau\nu}-\frac{\partial v^\nu}{\partial x^\tau}\xi_{\mu\nu}\right\}+v^\nu\left\{\frac{\partial}{\partial x^\mu}\left(\frac{\partial v_\tau}{\partial x^\nu}-\frac{\partial v_\nu}{\partial x^\tau}\right)-\frac{\partial}{\partial x^\tau}\left(\frac{\partial v_\mu}{\partial x^\nu}-\frac{\partial v_\nu}{\partial x^\mu}\right)\right\}, \tag{6·3}$$

and since
$$\frac{d}{ds}=v^\nu\frac{\partial}{\partial x^\nu}, \tag{6·4}$$

(6·2) and (6·3) yield

$$\frac{\partial g_\tau}{\partial x^\mu}-\frac{\partial g_\mu}{\partial x^\tau}=c\sqrt{(-g)}\left\{\frac{\partial v^\nu}{\partial x^\mu}\xi_{\tau\nu}-\frac{\partial v^\nu}{\partial x^\tau}\xi_{\mu\nu}\right\}+\frac{d}{ds}\{c\sqrt{(-g)}\,\xi_{\tau\mu}\},$$

or
$$\frac{d}{ds}\{c\sqrt{(-g)}\,\xi_{\tau\mu}\}=-c\sqrt{(-g)}\left\{\frac{\partial v^\nu}{\partial x^\mu}\xi_{\tau\nu}-\frac{\partial v^\nu}{\partial x^\tau}\xi_{\mu\nu}\right\}+\left(\frac{\partial g_\tau}{\partial x^\mu}-\frac{\partial g_\mu}{\partial x^\tau}\right). \tag{6·5}$$

From this equation, the formula for $d\xi_{\tau\mu}/ds$, the rate of change of the vorticity-tensor 'following the motion', may easily be obtained.

Proceeding again to the Newtonian approximation, we use (2·11), (3·2), (5·9) and (5·12) and we neglect terms of order higher than those in $1/c^2$. The equations (6·5) become

$$\frac{d}{dx^4}(\sqrt{\Delta}\,\xi_{\tau\mu})=-\sqrt{\Delta}\left\{\frac{\partial u^q}{\partial x^\mu}\xi_{\tau q}+\frac{\partial u^q}{\partial x^\tau}\xi_{q\mu}\right\}+\frac{\partial}{\partial x^\mu}\left\{\frac{1}{\rho}\frac{\partial p}{\partial x^\tau}-F_\tau\right\}-\frac{\partial}{\partial x^\tau}\left\{\frac{1}{\rho}\frac{\partial p}{\partial x^\mu}-F_\mu\right\}, \tag{6·6}$$

where, of course, the $\xi_{\mu\nu}$ are now given by (3·9) whilst d/dx^4 is defined by (5·14). These six formulae fall into two groups:

(i) *When either τ or μ is equal to* 4

The three formulae so obtained give essentially the rate of change of 'local' acceleration and could, for example, be used to calculate the quantities $\partial^2u^p/(\partial x^4)^2$ in problems of perturbed steady motions. But they must be employed with care for they involve terms in $\partial p/\partial x^4$ which arise through the slight modification of the equation of continuity implied in the definitions (4·3) and (4·5) of the energy-tensor. Indeed, in a practical case, it is better to calculate $\partial^2u^p/(\partial x^4)^2$ directly from the equations of motion (5·15) by partial differentiation with respect to x^4 (Sawyer 1947).

(ii) *When neither τ nor μ is equal to* 4

The corresponding three formulae now contain no terms to which exception can be taken and they yield the equations for the rate of change of vorticity 'following the motion'. To simplify them we again make use of the convention that *lmn* is to represent any cyclic permutation of 123 and we write

$$X_l=\frac{\partial F_m}{\partial x^n}-\frac{\partial F_n}{\partial x^m},\quad Y_l=\frac{\partial\rho^{-1}}{\partial x^m}\frac{\partial p}{\partial x^n}-\frac{\partial\rho^{-1}}{\partial x^n}\frac{\partial p}{\partial x^m}. \tag{6·7}$$

We write out (6·6) for the cases $\tau\mu$ equal to 23, 31 and 12, respectively, and use (3·13), (3·4) and (6·7) to obtain for the rate of change of vorticity $(\Omega_1, \Omega_2, \Omega_3)$,

$$\frac{d\Omega_l}{dx^4} = -\frac{d\omega_l}{dx^4} - \Theta(\Omega_l+\omega_l) + \sum_{p=1}^{3}(\Omega_p+\omega_p)\frac{\partial u^l}{\partial x^p} - \frac{X_l+Y_l}{\sqrt{\Delta}} \quad (l = 1, 2, 3). \tag{6·8}$$

We shall see in § 8 below that these are the generalizations of Helmholtz's equations. The vanishing of the right-hand sides of (6·8) is the condition that the vorticity be conserved 'following the motion'. This condition involves *kinematical* variables only, provided that

$$X_l+Y_l = 0 \quad (l = 1, 2, 3),$$

which would be the case, for example, if the force per unit volume (F_1, F_2, F_3) were derivable from a potential and if the variations of pressure and density were such that the quantities Y_l were zero (or negligibly small). The last-mentioned limitation would arise if the pressure were a function of the density alone.

7. HEAT TRANSFER

If our fluid is a gas we must take into account the changes of heat-content of unit mass during the motion. This presents us with a difficulty because, unlike the generalized mechanics we have hitherto drawn on, there is no four-dimensional theory of thermodynamics which has received universal assent (Tolman 1934). Since, however, our aim is to reach the Newtonian approximation, we can employ the following generalization in the case of a non-viscous perfect gas defined by the equation of state $p = R\rho T$. In terms of the fixed rectangular co-ordinates, we denote by dQ/dX^4 the rate of addition of heat to a fluid element 'following the motion'. Then if c_v is the specific heat at constant volume, k the conductivity and J the mechanical equivalent of heat, we have (Goldstein 1938)

$$J\frac{dQ}{dX^4} = Jc_v\frac{dT}{dX^4} + p\frac{d\rho^{-1}}{dX^4} - \frac{Jk}{\rho}\nabla^2 T, \tag{7·1}$$

where
$$\nabla^2 = \sum_{q=1}^{3}\frac{\partial^2}{(\partial X^q)^2}.$$

Let us now assume that T and Q, like p and ρ, are invariants and that k, c_v are constants. Then the four quantities $\partial T/\partial X^\mu$ $(\mu = 1, 2, 3, 4)$ are the components of a covariant four-vector. Using (2·13) we can therefore replace (7·1) by the four-dimensional tensor equation

$$J\frac{dQ}{ds} = Jc_v\frac{dT}{ds} + p\frac{d\rho^{-1}}{ds} + \frac{1}{c^2}\frac{Jk}{\rho}\Box T, \tag{7·2}$$

where
$$\Box = \frac{1}{\sqrt{(-g)}}\sum_{\mu=1}^{4}\frac{\partial}{\partial X^\mu}\left\{\sqrt{(-g)}\,g^{\mu\mu}\frac{\partial}{\partial X^\mu}\right\}. \tag{7·3}$$

In effect we have introduced a term $\dfrac{1}{c^2}\dfrac{\partial^2 T}{(\partial X^4)^2}$ into the right-hand side of (7·1) which will disappear in the Newtonian approximation provided that—as we know to be the case—temperature effects are propagated with a speed small compared with that of light. Hence, in terms of our moving curvilinear co-ordinate-system, the equation of heat-transfer becomes

$$J\frac{dQ}{ds} = Jc_v\frac{dT}{ds}+p\frac{d\rho^{-1}}{ds}+\frac{Jk}{\rho}\frac{1}{c^2}\frac{1}{\sqrt{(-g)}}\frac{\partial}{\partial x^\mu}\left\{\sqrt{(-g)}\,g^{\mu\nu}\frac{\partial T}{\partial x^\nu}\right\}. \tag{7·4}$$

We now consider the Newtonian approximation. Since

$$g^{4\nu}\frac{\partial T}{\partial x^\nu} = g^{44}\frac{\partial T}{\partial x^4}+g^{4q}\frac{\partial T}{\partial x^q}, \quad g^{q\nu}\frac{\partial T}{\partial x^\nu} = g^{q4}\frac{\partial T}{\partial x^4}+g^{qr}\frac{\partial T}{\partial x^r},$$

we observe that, by (2·12), the terms on the right-hand sides of these equations are of order unity except for the terms $g^{qr}\partial T/\partial x^r$, which are of order c^2. Hence using (2·11) and (2·12), $ds = dx^4$, and neglecting terms of order $1/c^2$, our Newtonian approximation to (7·4) reads

$$J\frac{dQ}{dx^4} = Jc_v\frac{dT}{dx^4}+p\frac{d\rho^{-1}}{dx^4}-\frac{Jk}{\rho}\frac{1}{\sqrt{\Delta}}\frac{\partial}{\partial x^q}\left\{\sqrt{\Delta}\,\gamma^{qr}\frac{\partial T}{\partial x^r}\right\}, \tag{7·5}$$

with d/dx^4 given by (5·14). This therefore is the equation of heat-transfer referred to a curvilinear co-ordinate-system in motion relative to the fixed rectangular system. We again note that the motion of the system will only affect the equation in cases when the γ^{qr} and $\sqrt{\Delta}$ are explicit functions of the time.

8. SPECIAL CASES

We illustrate the foregoing general theory by two examples which refer to the motion of a gas on a spherical rotating Earth.

(i) *Local rectangular co-ordinates*

It can be shown (McVittie 1948) that, if a tropospheric motion does not extend to greater distances than some 200 km. from a point O on the Earth's surface, a right-handed rectangular co-ordinate-system $Oxyz$ may be used of which the axes Ox, Oy lie in the tangent plane at O. If O is in North latitude ϕ and Ox is drawn to the East, the kinetic energy of unit mass of air is (McVittie 1948)

$$K = K_1-\omega^2K_2,$$

where

$$\left.\begin{aligned} 2K_1 &= \dot{x}^2+\dot{y}^2+\dot{z}^2+k\{\dot{x}(z+a)-x\dot{z}\}-l(\dot{x}y-x\dot{y}),\\ 2K_2 &= -[\{(z+a)\cos\phi-y\sin\phi\}^2+x^2],\\ k &= 2\omega\cos\phi, \quad l = 2\omega\sin\phi,\\ a &= \text{Earth's radius},\\ \omega &= \text{Earth's angular speed of rotation about its polar axis.}\end{aligned}\right\} \tag{8·1}$$

If in (2·7) we make the identifications

$$x^1 = x, \quad x^2 = y, \quad x^3 = z, \quad x^4 = t,$$

we have for the kinetic metric

$$ds^2 = \left(1+\frac{\omega^2 K_2}{c^2}\right) dt^2 - \frac{1}{c^2}[dx^2+dy^2+dz^2+\{k(z+a)-ly\}\,dx\,dt+lx\,dy\,dt-kx\,dz\,dt]. \tag{8·2}$$

We neglect terms in ω^2, as usual, and obtain

$$\left.\begin{gathered}\gamma_{44}=0,\quad \gamma_{14}=\gamma_{41}=\tfrac{1}{2}\{k(z+a)-ly\},\quad \gamma_{24}=\gamma_{42}=\tfrac{1}{2}lx,\quad \gamma_{34}=\gamma_{43}=-\tfrac{1}{2}kx,\\ \gamma_{11}=1,\quad \gamma_{22}=1,\quad \gamma_{33}=1,\quad \gamma_{pq}=0 \quad\text{when}\quad p\neq q,\\ \sqrt{\Delta}=1.\end{gathered}\right\} \tag{8·3}$$

If we also write $$u^1=u,\quad u^2=v,\quad u^3=w, \tag{8·4}$$

we obtain from (3·5) and (3·13) the following formulae for the divergence and vorticity

$$\Theta=\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}, \tag{8·5}$$

$$\left.\begin{gathered}\omega_1=0,\quad \omega_2=k,\quad \omega_3=l,\\ \Omega_1=\xi=\frac{\partial w}{\partial y}-\frac{\partial v}{\partial z},\quad \Omega_2=\eta=\frac{\partial u}{\partial z}-\frac{\partial w}{\partial x},\quad \Omega_3=\zeta=\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}.\end{gathered}\right\} \tag{8·6}$$

From (8·3) and (5·15), the reader may easily write down the usual equations of motion; we turn to the equations (6·8) expressing the rate of change of vorticity. Using (8·6) and remembering that dk/dt and dl/dt are both zero since we are restricted to a small area around O, we find

$$\left.\begin{aligned}\frac{d\xi}{dt}&=-\xi\Theta+\left\{\xi\frac{\partial u}{\partial x}+(k+\eta)\frac{\partial u}{\partial y}+(l+\zeta)\frac{\partial u}{\partial z}\right\}-X_1-Y_1,\\ \frac{d\eta}{dt}&=-(k+\eta)\Theta+\left\{\xi\frac{\partial v}{\partial x}+(k+\eta)\frac{\partial v}{\partial y}+(l+\zeta)\frac{\partial v}{\partial z}\right\}-X_2-Y_2,\\ \frac{d\zeta}{dt}&=-(l+\zeta)\Theta+\left\{\xi\frac{\partial w}{\partial x}+(k+\eta)\frac{\partial w}{\partial y}+(l+\zeta)\frac{\partial w}{\partial z}\right\}-X_3-Y_3,\end{aligned}\right\} \tag{8·7}$$

where the X_l, Y_l are given by (6·7) with $(x^1,x^2,x^3)\equiv(x,y,z)$. A particular case of the third of these formulae has been obtained by Petterssen (1936). He considers the case $w=0$ and takes $X_3=0$ so that the frictional forces have no x- or y-components. He also writes in vector notation

$$\mathbf{v}=v\mathbf{i}+u\mathbf{j},\quad \boldsymbol{\lambda}=2w\sin\phi\,.\,\mathbf{k},\quad \Theta=\operatorname{div}\mathbf{v},$$

$$\zeta\mathbf{k}=\operatorname{curl}\mathbf{v},\quad Y_3\mathbf{k}=\left(\frac{\partial\rho^{-1}}{\partial x}\frac{\partial p}{\partial y}-\frac{\partial\rho^{-1}}{\partial y}\frac{\partial p}{\partial x}\right)\mathbf{k}=-\nabla\rho^{-1}\times(-\nabla p).$$

Hence the third of the equations (8·7) becomes

$$\frac{d}{dt}(\operatorname{curl}\mathbf{v})=-(\boldsymbol{\lambda}+\operatorname{curl}\mathbf{v})\operatorname{div}\mathbf{v}+\nabla\rho^{-1}\times(-\nabla p),$$

which is Petterssen's formula. We note, however, that by itself, it tells us little about the rate of change of the vorticity because, under the conditions assumed by him, the first two equations (8·7) are not identically satisfied.

If the motion of the air which we are studying takes place in the immediate vicinity of O (as in aerodynamical problems) we can neglect the terms in k and l in (8·7) and consequently draw the axis Ox in any direction relatively to the east-west line through O. Formulae (8·7) then give us the expressions (referred to 'fixed' rectangular axes) for the rate of change of vorticity with any relationship between p and ρ, and any external force per unit volume. To obtain the classical equations due to Helmholtz (Lamb 1930) we proceed thus:

The equation of continuity may be written, in our case,

$$\frac{1}{\rho}\frac{d\rho}{dt}+\Theta = 0,$$

and if, as Helmholtz assumes, p is a function of ρ alone,* then by (6·7)

$$Y_l = 0 \quad (l = 1, 2, 3).$$

If, moreover, we assume that the external force per unit volume is derived from a potential (i.e. no frictional forces act), we have

$$X_l = 0 \quad (l = 1, 2, 3).$$

Hence the first of the equations (8·7) becomes

$$\frac{d\xi}{dt}-\frac{\xi}{\rho}\frac{d\rho}{dt} = \xi\frac{\partial u}{\partial x}+\eta\frac{\partial u}{\partial y}+\zeta\frac{\partial u}{\partial z}$$

or

$$\frac{d(\xi/\rho)}{dt} = \frac{1}{\rho}\left(\xi\frac{\partial u}{\partial x}+\eta\frac{\partial u}{\partial y}+\zeta\frac{\partial u}{\partial z}\right),$$

with corresponding equations for $\frac{d(\eta/\rho)}{dt}$, $\frac{d(\zeta/\rho)}{dt}$, as found by Helmholtz.

(ii) *Local cylindrical polars*

If in (8·2) we replace (x, y) by (r, θ), where $x = r\sin\theta$, $y = -r\cos\theta$, so that $\theta = 0$ is the negative Oy axis, the kinetic metric takes the form (again neglecting the term in ω^2),

$$ds^2 = dt^2-\frac{1}{c^2}[dr^2+r^2d\theta^2+dz^2+k(z+a)\sin\theta\, dr\, dt + \{k(z+a)r\cos\theta+lr^2\}\, d\theta\, dt - kr\sin\theta\, dz\, dt]. \tag{8·8}$$

We now make the identifications $x^1 = r$, $x^2 = \theta$, $x^3 = z$, $x^4 = t$ and we have

$$\left.\begin{array}{c} \gamma_{44} = 0, \quad \gamma_{14} = \gamma_{41} = \tfrac{1}{2}k(z+a)\sin\theta, \quad \gamma_{24} = \gamma_{42} = \tfrac{1}{2}\{k(z+a)r\cos\theta+lr^2\}, \\ \gamma_{34} = \gamma_{43} = -\tfrac{1}{2}kr\sin\theta, \\ \gamma_{11} = 1, \quad \gamma_{22} = r^2, \quad \gamma_{33} = 1, \quad \gamma_{pq} = 0 \quad \text{when } p \neq q, \quad \sqrt{\Delta} = r. \end{array}\right\} \tag{8·9}$$

We also introduce the linear components of velocity, viz:

$$U = u^1, \quad V = ru^2, \quad W = u^3.$$

---

* This restriction is removed in the work of I. Silberstein, *Bull. Int. Acad. Sci.*, Cracovie, *C.R.* 1896, p. 280.

The divergence and *absolute* vorticity become by (3·4), (3·13) and (8·9),

$$\left.\begin{aligned}\Theta &= \frac{\partial U}{\partial r}+\frac{U}{r}+\frac{1}{r}\frac{\partial V}{\partial \theta}+\frac{\partial W}{\partial z},\\ \xi_1 &= -k\cos\theta-\frac{1}{r}\left(r\frac{\partial V}{\partial z}-\frac{\partial W}{\partial \theta}\right),\\ \xi_2 &= +\frac{k\sin\theta}{r}-\frac{1}{r}\left(\frac{\partial W}{\partial r}-\frac{\partial U}{\partial z}\right),\\ \xi_3 &= l-\frac{1}{r}\left\{\frac{\partial U}{\partial \theta}-\frac{\partial (rV)}{\partial r}\right\}.\end{aligned}\right\} \tag{8·10}$$

We consider Sawyer's theory (1947) of the development of tropical cyclones in the light of equations (6·8). Sawyer discusses the small perturbations of a symmetrical circular motion, produced by a slight change of pressure-gradient. This basic flow is defined by

$$\left.\begin{aligned}U = 0, \quad V = v_0(r,z), \quad W = 0,\\ lv_0+\frac{v_0^2}{r} = \frac{1}{\rho_0}\frac{\partial p_0}{\partial r},\end{aligned}\right\} \tag{8·11}$$

where p_0, ρ_0 are also functions of r and z alone. The stability of the basic flow is shown to depend on the sign of the quantity

$$B = \left(l+\frac{2v_0}{r}\right)\left(l+\frac{\partial v_0}{\partial r}+\frac{v_0}{r}\right), \tag{8·12}$$

which must be negative for *instability* and consequent development of the tropical cyclone. Using (8·10), (8·11) and (6·7) we obtain for the basic flow

$$\left.\begin{aligned}\xi_1 = -k\cos\theta-\frac{\partial v_0}{\partial z}, \quad \xi_2 = \frac{k\sin\theta}{r}, \quad \xi_3 = l+\frac{\partial v_0}{\partial r}+\frac{v_0}{r},\\ \Theta = 0, \quad X_1 = X_2 = X_3 = 0,\\ Y_1 = 0, \quad Y_2 = \frac{\partial \rho_0^{-1}}{\partial z}\frac{\partial p_0}{\partial r}-\frac{\partial \rho_0^{-1}}{\partial r}\frac{\partial p_0}{\partial z}, \quad Y_3 = 0.\end{aligned}\right\} \tag{8·13}$$

Hence by (8·4) and (8·13) the three equations (6·8) may be written

$$\begin{aligned}&\frac{d\xi_1}{dx^4} = 0,\\ &\frac{d\xi_2}{dx^4} = \xi_1\frac{\partial(v_0/r)}{\partial r}+\xi_3\frac{\partial(v_0/r)}{\partial z}-\frac{Y_2}{r},\\ &\frac{d\xi_3}{dx^4} = 0.\end{aligned}$$

Since our variables depend on r, z alone, we find, on evaluating the left-hand sides of these equations using (8·13), that the third equation alone is an identity. The other two are

$$\begin{aligned}&k\frac{v_0}{r}\sin\theta = 0,\\ &k\frac{v_0}{r^2}\cos\theta = -\left(k\cos\theta+\frac{\partial v_0}{\partial z}\right)\left(\frac{1}{r}\frac{\partial v_0}{\partial r}-\frac{v_0}{r^2}\right)+\left(l+\frac{\partial v_0}{\partial r}+\frac{v_0}{r}\right)\frac{1}{r}\frac{\partial v_0}{\partial z}-\frac{Y_2}{r},\end{aligned}$$

or, on re-arranging the terms in the second equation,

$$k\frac{v_0}{r}\sin\theta = 0, \tag{8·14}$$

$$\frac{k\cos\theta}{r}\left(\frac{\partial v_0}{\partial r}\right) = \frac{1}{r}\frac{\partial v_0}{\partial z}\left(l+\frac{2v_0}{r}\right)-\frac{Y_2}{r}. \tag{8·15}$$

These two equations contain obvious inconsistencies; hence Sawyer's basic flow is possible only if we regard $\frac{kv_0}{r}$ and $\frac{k}{r}\frac{\partial v_0}{\partial r}$ as negligibly small, a somewhat precarious assumption when we remember that in equatorial regions, k approaches its maximum value of 2ω. However, if we do make this assumption, we obtain from (8·15)

$$\frac{\partial v_0}{\partial z}\left(l+\frac{2v_0}{r}\right) = Y_2 = \frac{\partial \rho_0^{-1}}{\partial z}\frac{\partial p_0}{\partial r}-\frac{\partial \rho_0^{-1}}{\partial r}\frac{\partial p_0}{\partial z}. \tag{8·16}$$

If we now assume that $\partial\rho_0^{-1}/\partial r$ is negligibly small, (8·16) and (8·12) yield

$$B = -\left\{\frac{1}{\rho_0^2}\frac{\partial \rho_0}{\partial z}\frac{\partial p_0}{\partial r}\bigg/\frac{\partial v_0}{\partial z}\right\}\left\{l+\frac{\partial v_0}{\partial r}+\frac{v_0}{r}\right\}.$$

But, in general in the atmosphere, $\partial\rho_0/\partial z < 0$, hence B is negative if either

(*a*) $l+\frac{\partial v_0}{\partial r}+\frac{v_0}{r} > 0$ (i.e. $\xi_3 > 0$) and the pressure-gradient of the basic flow, $\frac{\partial p_0}{\partial r}$, has the *opposite* sign to $\partial v_0/\partial z$; or

(*b*) $l+\frac{\partial v_0}{\partial r}+\frac{v_0}{r} < 0$ (i.e. $\xi_3 < 0$) and the pressure-gradient has the *same* sign as $\partial v_0/\partial z$.

These expressions for Sawyer's criterion of instability are perhaps of a somewhat more physical character than those he himself adopts.

References

Eddington, A. S. 1924 *Mathematical theory of relativity*. 2nd ed. p. 111. Cambridge University Press.

Goldstein, S. 1938 *Modern developments in fluid dynamics*. Vol. II, § XIV.

Lamb, H. 1930 *Hydrodynamics*. 5th ed. p. 188. Cambridge University Press.

McConnell, A. J. 1931 *Applications of the absolute differential calculus*, pp. 277, 146, eqn (14). London: Blackie and Son, Ltd.

McVittie, G. C. 1948 *Quart. J. Mech. App. Math.* **1**, 174, §§ 4, 7.

Petterssen, S. 1936 *Geofys. Pub. Oslo*. **11**, no. 6, 23.

Sawyer, J. S. 1947 *Quart. J. Roy. Met. Soc.* **73**, 101.

Tolman, R. C. 1934 *Relativity, thermodynamics and cosmology*. Oxford University Press.

Two-dimensional fluid motion referred to a network of orthogonal curves

By G. C. McVittie, *Queen Mary College, London*

(*Communicated by G. Temple, F.R.S.—Received* 3 *August* 1948)

A theory already developed is applied to the case of two-dimensional motion parallel at each point of space to some member, Σ, of a one-parameter family of surfaces, the co-ordinate-system being a network of orthogonal curves drawn on Σ. The geodesic curvatures of the orthogonal curves and their relationship to the Gaussian curvature of Σ are worked out. The equations of motion and of continuity are expressed in terms of the geodesic curvatures. Meyer's aerodynamical equations are derived as particular cases when the network is fixed in space and the surfaces are all planes. A formula for a large-scale gradient wind is also obtained as an example of the use of a moving network drawn on a sphere.

1. Introduction

In a previous investigation (McVittie 1949) there has been developed a method for establishing the equations of classical hydrodynamics relative to a curvilinear system of co-ordinates (x^1, x^2, x^3) in motion with respect to a fixed rectangular system (X^1, X^2, X^3). The fundamental entity is the kinetic metric*

$$\left.\begin{aligned} ds^2 &= \left(1-\frac{\gamma_{44}}{c^2}\right)(dx^4)^2 - \frac{1}{c^2}\{\gamma_{pq}\,dx^p\,dx^q + 2\gamma_{p4}\,dx^p\,dx^4\}, \\ &= g_{\mu\nu}\,dx^\mu\,dx^\nu, \end{aligned}\right\} \tag{1·1}$$

where the $\gamma_{\mu\nu}$ are the coefficients in the quadratic form, K, which gives the kinetic energy of unit mass of fluid, and, for the sake of uniformity, the (absolute) Newtonian time is denoted by x^4, whilst c is the velocity of light. It was then shown that the Newtonian equations of motion and of continuity, respectively, were (no summation convention)

$$\frac{d}{dx^4}(\gamma_{qq}u^q) - \frac{1}{2}\sum_{p=1}^{3}\frac{\partial\gamma_{pp}}{\partial x^q}(u^p)^2 - \left(\frac{1}{2}\frac{\partial\gamma_{44}}{\partial x^q} - \frac{\partial\gamma_{q4}}{\partial x^4}\right) - \sum_{p=1}^{3}\left(\frac{\partial\gamma_{p4}}{\partial x^q} - \frac{\partial\gamma_{q4}}{\partial x^p}\right)u^p = -\frac{1}{\rho}\frac{\partial p}{\partial x^q} + F_q, \tag{1·2}$$

$$\frac{\partial(\sqrt{\Delta}\rho)}{\partial x^4} + \sum_{p=1}^{3}\frac{\partial(\sqrt{\Delta}\rho u^p)}{\partial x^p} = 0, \tag{1·3}$$

where $\dfrac{d}{dx^4} = \dfrac{\partial}{\partial x^4} + \sum\limits_{p=1}^{3} u^p \dfrac{\partial}{\partial x^p}$, the u^p $(p = 1, 2, 3)$ are the velocity-components of the fluid relative to the moving axes, F_q $(q = 1, 2, 3)$ are the components of force per unit volume and Δ is the determinant of the γ_{pq}. In (1·2) it is assumed that the co-ordinate-system is spatially orthogonal, i.e. that $\gamma_{pq} = 0$ when $p \neq q$.

In certain problems of aerodynamics and of dynamical meteorology the motion of the fluid at each point of space is, in the main, parallel to some member of a one-parameter family of planes or surfaces. In such cases it is convenient to refer the motion to a network of orthogonal curves drawn in this plane or surface and the geodesic curvatures of these curves then figure with advantage in the equations of motion and of continuity. We consider the general theory of this kind of representation.

---

* Greek indices can take the values 1 to 4, Roman indices the values 1 to 3. The summation convention is used throughout, except where otherwise stated.

2. GEOMETRICAL THEORY

(i) Consider first the four-dimensional manifold (1·1) in which (x^1, x^2, x^3) is a co-ordinate-system arbitrarily chosen except that $\gamma_{pq} = 0$ when $p \neq q$. At every instant x^4 of Newtonian time, its 'spatial section' is a three-space with metric

$$ds_0^2 = \sum_{p=1}^{3} \gamma_{pp}(dx^p)^2, \tag{2·1}$$

whether the co-ordinate-system is in motion or at rest relative to the fixed rectangular axes. In general the γ_{pp} are functions of (x^1, x^2, x^3) and may also involve the particular value of x^4 at which the 'spatial section' of the four-dimensional manifold has been taken. Let us also for definiteness take the surface $x^3 = 0$ as that on which the co-ordinate-system (x^1, x^2) has been laid out and denote this surface by Σ_0. Suppose now that on Σ_0 we trace two systems of orthogonal curves whose equations are

$$\alpha^1 = \phi_1(x^1, x^2), \quad \alpha^2 = \phi_2(x^1, x^2).$$

The solutions of these equations would give

$$x^1 = \chi_1(\alpha^1, \alpha^2), \quad x^2 = \chi_2(\alpha^1, \alpha^2), \tag{2·2}$$

and by taking their differentials we could express (2·1) as

$$ds_0^2 = \sum_{p=1}^{2} a_{pp}(d\alpha^p)^2 + a_{33}(dx^3)^2, \tag{2·3}$$

where a_{33} is the result of replacing x^1, x^2 by α^1, α^2 in γ_{33}. From the four-dimensional point of view all that we have done, however, is to apply a transformation

$$x^1 = \chi_1(\alpha^1, \alpha^2), \quad x^2 = \chi_2(\alpha^1, \alpha^2), \quad x^3 = \alpha^3, \quad x^4 = \alpha^4 \tag{2·4}$$

to (1·1) and so to transform it into

$$\left.\begin{aligned} ds^2 &= \left(1 - \frac{a_{44}}{c^2}\right)(d\alpha^4)^2 - \frac{1}{c^2}\left\{\sum_{p=1}^{3} a_{pp}(d\alpha^p)^2 + 2a_{p4}\,d\alpha^p\,d\alpha^4\right\}, \\ &= \bar{g}_{\mu\nu}\,d\alpha^\mu\,d\alpha^\nu, \end{aligned}\right\} \tag{2·5}$$

where the $a_{\mu\nu}$ are obtained from the $\gamma_{\mu\nu}$ by the rule for transformation of the $g_{\mu\nu}$ under (2·4), viz.

$$\bar{g}_{\mu\nu} = g_{\lambda\tau}\frac{\partial x^\lambda}{\partial \alpha^\mu}\frac{\partial x^\tau}{\partial \alpha^\nu}. \tag{2·6}$$

(ii) The next step is to consider the intrinsic geometry of the α^1- and α^2-curves drawn on Σ_0. The metric of this surface at the instant x^4 is by (2·3)

$$d\sigma^2 = a_{11}(d\alpha^1)^2 + a_{22}(d\alpha^2)^2,$$

where we must remember that we have replaced α^3 by zero in the functions a_{11} and a_{22}. We introduce the convention that the indices i and j can take the values 1 and 2 only and we write

$$d\sigma^2 = a_{ij}d\alpha^i d\alpha^j \quad (a_{12} = 0). \tag{2·7}$$

It is required to express the *geodesic* or *surface* curvatures of the α^1- and α^2-curves in terms of the a_{ij} and their derivatives. The signs to be given to the curvatures are a matter of convention and writers differ in this matter. We adopt the convention of Meyer (1948) and take as our standard type of intersection that indicated at the point P in figure 1. The arrows on the curves PS, PR indicate the directions of increasing α^1, α^2 respectively. The tangent vectors at P, denoted by $\mathbf{t}_{(1)}$, $\mathbf{t}_{(2)}$ are taken positive in the directions of increasing α^1, α^2, respectively. The positive direction of the normal vector $\mathbf{n}_{(1)}$ to the α^1-curve at P is shown by the direction of the dotted arrow and similarly for the normal vector $\mathbf{n}_{(2)}$ to the α^2-curve. We take these four vectors to be unit surface vectors in Σ_0. Using (2·7) we have for the (surface) components of the tangent vectors

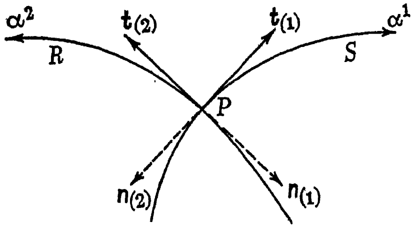


FIGURE 1

$$\left.\begin{aligned} \mathbf{t}_{(1)} &= (\lambda^1_{(1)}, \lambda^2_{(1)}) = \left(\frac{1}{\sqrt{a_{11}}}, 0\right), \\ \mathbf{t}_{(2)} &= (\lambda^1_{(2)}, \lambda^2_{(2)}) = \left(0, \frac{1}{\sqrt{a_{22}}}\right). \end{aligned}\right\} \tag{2·8}$$

Since $\mathbf{t}_{(1)}$ is a unit vector, $a_{ij}\lambda^i_{(1)}\lambda^j_{(1)} = 1$; and hence, taking the intrinsic derivative (McConnell 1931) with respect to (2·7) of both sides of this equation, we obtain

$$a_{ij}\frac{\delta\lambda^i_{(1)}}{\delta\sigma}\lambda^j_{(1)} = 0,$$

where we denote the intrinsic derivative by $\delta/\delta\sigma$. Hence, since $\delta\sigma$ represents a displacement along the α^1-curve, the last equation means (McConnell 1931) that the vector $\delta\lambda^i_{(1)}/\delta\sigma$ $(i = 1, 2)$ is perpendicular to the tangent vector $\mathbf{t}_{(1)}$, i.e. it is in the direction of the unit vector $\mathbf{n}_{(1)}$. If therefore we write $\mathbf{n}_{(1)} = (\mu^1_{(1)}, \mu^2_{(1)})$ and denote the geodesic curvature of the α^1-curve by κ_1, we have

$$\frac{\delta\lambda^i_{(1)}}{\delta\sigma} = \kappa_1\mu^i_{(1)}. \tag{2·9}$$

Similarly for the α^2-curves

$$\frac{\delta\lambda^i_{(2)}}{\delta\sigma} = \kappa_2\mu^i_{(2)}, \tag{2·10}$$

where $\mathbf{n}_{(2)} = (\mu^1_{(2)}, \mu^2_{(2)})$ and κ_2 is the geodesic curvature of the α^2-curve. But now with the sign convention of figure 1, we have

$$\mathbf{n}_{(1)} = -\mathbf{t}_{(2)}, \quad \mathbf{n}_{(2)} = -\mathbf{t}_{(1)}. \tag{2·11}$$

Hence writing out the intrinsic derivatives in terms of the Christoffel symbols and using (2·11), the equations (2·9) become

$$\left.\begin{aligned} \frac{\delta\lambda^1_{(1)}}{\delta\sigma} &= \frac{d\lambda^1_{(1)}}{d\sigma} + \begin{Bmatrix} 1 \\ ij \end{Bmatrix} \lambda^i_{(1)}\lambda^j_{(1)} = -\kappa_1\lambda^1_{(2)}, \\ \frac{\delta\lambda^2_{(1)}}{\delta\sigma} &= \frac{d\lambda^2_{(1)}}{d\sigma} + \begin{Bmatrix} 2 \\ ij \end{Bmatrix} \lambda^i_{(1)}\lambda^j_{(1)} = -\kappa_1\lambda^2_{(2)}. \end{aligned}\right\} \tag{2·12}$$

But on referring to (2·8), we observe that the first of the equations (2·12) is an identity; the left-hand side being zero because the intrinsic derivative of $(a_{11})^{-\frac{1}{2}}$ is zero and the right-hand side, because $\lambda^1_{(2)}$ is zero. The second equation (2·12) reduces to

$$\begin{Bmatrix} 2 \\ 11 \end{Bmatrix} (\lambda^1_{(1)})^2 = -\frac{\kappa_1}{\sqrt{a_{22}}},$$

whence on evaluating the Christoffel symbol, we obtain

$$-\left(\frac{1}{2}\frac{1}{a_{22}}\frac{\partial a_{11}}{\partial \alpha^2}\right)\frac{1}{a_{11}} = -\frac{\kappa_1}{\sqrt{a_{22}}}.$$

Hence finally,
$$\kappa_1 = \frac{1}{2}\frac{1}{a_{11}\sqrt{a_{22}}}\frac{\partial a_{11}}{\partial \alpha^2} = \frac{1}{\sqrt{(a_{11}a_{22})}}\frac{\partial\sqrt{a_{11}}}{\partial \alpha^2}; \tag{2·13}$$

and similarly
$$\kappa_2 = \frac{1}{2}\frac{1}{a_{22}\sqrt{a_{11}}}\frac{\partial a_{22}}{\partial \alpha^1} = \frac{1}{\sqrt{(a_{11}a_{22})}}\frac{\partial\sqrt{a_{22}}}{\partial \alpha^1}. \tag{2·14}$$

These are the required expressions for the geodesic curvatures of the orthogonal co-ordinate curves in the surface Σ_0. The curvatures possess an important property: if the α^1-curves, say, are *geodesics* of the surface Σ_0, then $\kappa_1 = 0$, a result which follows from (2·9) and the definition of a geodesic as a curve whose tangent has a zero intrinsic derivative (Eisenhart 1940).

(iii) Lastly, we consider the Gaussian curvature of the surface Σ_0 itself. For the metric (2·7) we have

$$||a_{ij}|| = a_{11}a_{22}, \quad a^{11} = \frac{1}{a_{11}}, \quad a^{22} = \frac{1}{a_{22}}, \tag{2·15}$$

and the non-zero component R_{1212} of the curvature-tensor (Eisenhart 1940) is

$$R_{1212} = -\frac{1}{2}\left\{\frac{\partial^2 a_{11}}{(\partial\alpha^2)^2} + \frac{\partial^2 a_{22}}{(\partial\alpha^1)^2}\right\} + \frac{1}{4}\left\{\frac{1}{a_{11}}\left(\frac{\partial a_{11}}{\partial\alpha^2}\right)^2 + \frac{1}{a_{22}}\left(\frac{\partial a_{22}}{\partial\alpha^1}\right)^2 + \frac{1}{a_{11}}\frac{\partial a_{11}}{\partial\alpha^1}\frac{\partial a_{22}}{\partial\alpha^1} + \frac{1}{a_{22}}\frac{\partial a_{22}}{\partial\alpha^2}\frac{\partial a_{11}}{\partial\alpha^2}\right\}. \tag{2·16}$$

The Gaussian curvature is defined by

$$K_0 = R_{1212}/(a_{11}a_{22}); \tag{2·17}$$

hence substituting for $\partial a_{11}/\partial\alpha^2$ and $\partial a_{22}/\partial\alpha^1$ from (2·13) and (2·14), we get, after some calculation,

$$K_0 = -\left\{\frac{\partial\kappa_2}{\sqrt{a_{11}}\,\partial\alpha^1} + \frac{\partial\kappa_1}{\sqrt{a_{22}}\,\partial\alpha^2} + (\kappa_2)^2 + (\kappa_1)^2\right\}. \tag{2·18}$$

This is the form of Laguerre's formula for the sign-convention of figure 1, the two sets of curves being orthogonal.

3. THE EQUATIONS OF MOTION AND OF CONTINUITY

We can generalize the theory of § 2 by supposing that the transformation (2·4) is valid for every member Σ of the one-parameter family of surfaces $x^3 = \alpha^3 =$ *constant* and for every instant x^4. The relationship of the members of the family to Σ_0 will depend on whether x^3 is a linear or an angular co-ordinate, a point illustrated in § 4. Since the $\gamma_{\mu\nu}$ of (1·1) or the $a_{\mu\nu}$ of (2·5) are functions of x^3 ($= \alpha^3$), the identity of the functions χ_1 and χ_2 from surface to surface does not imply that the kinetic energy of unit mass is the same on each surface Σ for identical values of α^1 and α^2. With these limitations we can regard κ_1, κ_2 and K_0 as being functions of α^3 as well as of α^1 and α^2; and, of course, should the $\gamma_{\mu\nu}$ be functions of α^4 also, we can suppose that the three curvatures vary with the time as well.

We introduce the three functions, h_p, where,

$$a_{pp} = h_p^2 \quad (p = 1, 2, 3), \tag{3·1}$$

and also the physical components of velocity V^p, where

$$\frac{d\alpha^p}{d\alpha^4} = u^p = V^p/h_p. \tag{3·2}$$

In the equations (1·2) we write

$$x^\mu = \alpha^\mu, \quad \gamma_{\mu\nu} = a_{\mu\nu} \quad (\mu, \nu = 1, 2, 3, 4),$$

$$\frac{d}{dx^4} = \frac{\partial}{\partial\alpha^4} + \sum_{p=1}^{3} \frac{V^p}{h_p}\frac{\partial}{\partial\alpha^p},$$

and we abandon the summation convention. The equations become

$$\frac{dV^q}{d\alpha^4} + \frac{V^q}{h_q}\frac{\partial h_q}{\partial\alpha^4} + \sum_{p=1}^{3}\frac{V^p}{h_p h_q}\left(\frac{\partial h_q}{\partial\alpha^p}V^q - \frac{\partial h_p}{\partial\alpha^q}V^p\right) - \frac{1}{h_q}\left(\frac{1}{2}\frac{\partial a_{44}}{\partial\alpha^q} - \frac{\partial a_{q4}}{\partial\alpha^4}\right) - \sum_{p=1}^{3}\frac{V^p}{h_p h_q}\left(\frac{\partial a_{p4}}{\partial\alpha^q} - \frac{\partial a_{q4}}{\partial\alpha^p}\right)$$
$$= \frac{1}{h_q}\left(-\frac{1}{\rho}\frac{\partial p}{\partial\alpha^q} + F_q\right). \tag{3·4}$$

In most practical cases, the h_p are not explicit functions of the time. Now if $V^3 \neq 0$, there is no advantage in replacing h_1, h_2 in (3·4) by the geodesic curvatures; if, however, V^3 is either zero or negligibly small, the equations of motion become, writing $\alpha^4 = t$, and using (2·13), (2·14) and (3·1),

$$\frac{dV^1}{dt} + \kappa_1 V^1 V^2 - \kappa_2 (V^2)^2 - \frac{1}{h_1}\left(\frac{1}{2}\frac{\partial a_{44}}{\partial\alpha^1} - \frac{\partial a_{14}}{\partial t}\right) - \frac{V^2}{h_1 h_2}\left(\frac{\partial a_{24}}{\partial\alpha^1} - \frac{\partial a_{14}}{\partial\alpha^2}\right) = \frac{1}{h_1}\left(-\frac{1}{\rho}\frac{\partial p}{\partial\alpha^1} + F_1\right), \tag{3·5}$$

$$\frac{dV^2}{dt}-\kappa_1(V^1)^2+\kappa_2 V^1V^2-\frac{1}{h_2}\left(\frac{1}{2}\frac{\partial a_{44}}{\partial\alpha^2}-\frac{\partial a_{24}}{\partial t}\right)-\frac{V^1}{h_1h_2}\left(\frac{\partial a_{14}}{\partial\alpha^2}-\frac{\partial a_{24}}{\partial\alpha^1}\right)=\frac{1}{h_2}\left(-\frac{1}{\rho}\frac{\partial p}{\partial\alpha^2}+F_2\right), \tag{3·6}$$

$$-\left\{\frac{1}{h_1}\frac{\partial h_1}{\partial\alpha^3}(V^1)^2+\frac{1}{h_2}\frac{\partial h_2}{\partial\alpha^3}(V^2)^2\right\}-\left(\frac{1}{2}\frac{\partial a_{44}}{\partial\alpha^3}-\frac{\partial a_{34}}{\partial t}\right)$$
$$-\frac{V^1}{h_1}\left(\frac{\partial a_{14}}{\partial\alpha^3}-\frac{\partial a_{34}}{\partial\alpha^1}\right)-\frac{V^2}{h_2}\left(\frac{\partial a_{24}}{\partial\alpha^3}-\frac{\partial a_{34}}{\partial\alpha^2}\right)=-\frac{1}{\rho}\frac{\partial p}{\partial\alpha^3}+F_3. \tag{3·7}$$

The equation of continuity in the case when V^3 is zero or negligibly small and none of the h_p contain the time t explicitly, becomes on writing $\sqrt{\Delta}=h_1h_2h_3$ in (1·3) and using (2·13), (2·14), (3·1) and (3·2),

$$\frac{\partial\rho}{\partial t}+\frac{1}{h_1}\frac{\partial\rho V^1}{\partial\alpha^1}+\frac{1}{h_2}\frac{\partial\rho V^2}{\partial\alpha^2}+\rho(\kappa_1V^2+\kappa_2V^1)+\frac{\rho}{h_3}\left(\frac{V^1}{h_1}\frac{\partial h_3}{\partial\alpha^1}+\frac{V^2}{h^2}\frac{\partial h_3}{\partial\alpha^2}\right)=0. \tag{3·8}$$

Lastly, by (2·18) and (3·1), the Gaussian curvature of any surface Σ is

$$K_0=-\left\{\frac{\partial\kappa_2}{h_1\partial\alpha^1}+\frac{\partial\kappa_1}{h_2\partial\alpha^2}+(\kappa_1)^2+(\kappa_2)^2\right\}. \tag{3·9}$$

4. Meyer's equations in aerodynamics

We discuss Meyer's equations of motion and of continuity (Meyer 1948) for two-dimensional flow in the light of § 3. We assume that the co-ordinate-system $(\alpha^1,\alpha^2,\alpha^3)$ is at rest with respect to fixed rectangular axes so that, in (3·5) to (3·7) $a_{\mu4}=0$ $(\mu=1,2,3,4)$. We also assume that the flow is steady, so that the $\partial V^p/\partial t$ are zero, that ρ, p do not involve the time explicitly, that there are no external forces acting $(F_p=0)$, and that the motion depends on two independent variables only, so that the functions h_p involve α^1, α^2 alone. With these definitions of the nature of the co-ordinate-system and of the flow, equation (3·7) shows that p is also a function of α^1, α^2 alone. We then write

$$\alpha^1=\alpha,\quad \alpha^2=\beta,\quad V^1=v_\alpha,\quad V^2=v_\beta,\quad \kappa_1=\kappa_\alpha,\quad \kappa_2=\kappa_\beta, \tag{4·1}$$

and (3·5), (3·6) become, respectively,

$$\left.\begin{aligned} v_\alpha\frac{\partial v_\alpha}{h_\alpha\partial\alpha}+v_\beta\frac{\partial v_\alpha}{h_\beta\partial\beta}+\kappa_\alpha v_\alpha v_\beta-\kappa_\beta v_\beta^2&=-\frac{1}{\rho}\frac{\partial p}{h_\alpha\partial\alpha},\\ v_\alpha\frac{\partial v_\beta}{h_\alpha\partial\alpha}+v_\beta\frac{\partial v_\beta}{h_\beta\partial\beta}-\kappa_\alpha v_\alpha^2+\kappa_\beta v_\alpha v_\beta&=-\frac{1}{\rho}\frac{\partial p}{h_\beta\partial\beta},\end{aligned}\right\} \tag{4·2}$$

which are Meyer's equations (8) and (9). Equation (3·8) becomes

$$\frac{\partial\rho v_\alpha}{h_\alpha\partial\alpha}+\frac{\partial\rho v_\beta}{h_\beta\partial\beta}+\rho(\kappa_\alpha v_\beta+\kappa_\beta v_\alpha)=-\rho\left(v_\alpha\frac{\partial\log h_3}{h_\alpha\partial\alpha}+v_\beta\frac{\partial\log h_3}{h_\beta\partial\beta}\right). \tag{4·3}$$

Meyer also imposes the condition

$$\frac{\partial\kappa_\beta}{h_\alpha\partial\alpha}+\frac{\partial\kappa_\alpha}{h_\beta\partial\beta}+(\kappa_\alpha)^2+(\kappa_\beta)^2=0,$$

which by (3·9) shows that the surfaces Σ are Euclidean planes. This means that, in terms of the fundamental co-ordinate-system (x^1, x^2, x^3) used in (2·1), the three-space has the metric

$$ds_0^2 = (dx^1)^2 + (dx^2)^2 + \gamma_{33}(dx^3)^2, \tag{4·4}$$

where γ_{33} is a function of x^1, x^2 alone. Since a_{33} $(= h_3^2)$ is simply the result of replacing, in γ_{33}, x^1, x^2 by α^1, α^2 through (2·2), the right-hand side of (4·3) depends ultimately on the way in which the kinetic energy of unit mass of fluid is assumed to vary as we pass for a given (α^1, α^2) from one plane Σ to another. Meyer's forms (Meyer 1948, eqn. (10)) for the equation of continuity may be obtained thus:

(i) *Plane flow.* Here we take x^1, x^2 to be rectangular co-ordinates in the plane of flow whilst x^3 is measured along the normal to all the planes Σ which are therefore parallel. Then by choosing $\gamma_{33} = h_3^2 = 1$ we find from (4·3), Meyer's equation of continuity for his case $j = 0$.

(ii) *Axially symmetrical flow.* Here we choose x^2 to be a linear co-ordinate measured along the axis of symmetry, x^1 to be a linear co-ordinate perpendicular to the axis and x^3 to be an angular co-ordinate giving orientation about the axis. The planes Σ now pass through the axis of symmetry. If we also choose $\gamma_{33} = h_3^2 = (x^1)^2$ in (4·4) and refer to Meyer's figure 1, we have $x^1 \equiv r$ and

$$\partial h_3/h_\alpha \partial\alpha \equiv \partial r/h_\alpha \partial\alpha = \sin\epsilon, \quad \partial h_3/h_\beta \partial\beta \equiv \partial r/h_\beta \partial\beta = \cos\epsilon,$$

so that the right-hand side of (4·3) reduces to Meyer's equation of continuity for his case $j = 1$.

5. GRADIENT WIND IN DYNAMICAL METEOROLOGY

An illustration of the manipulation of equations (3·5) to (3·8) for a moving co-ordinate-system is provided by the theory of the gradient wind, a wind which blows along the isobars. We consider a large-scale motion of this kind, i.e. a wind which belongs to a disturbance of the size of a cyclone of temperate latitudes with a diameter of up to 2000 km. It can be shown (McVittie 1948) that the equations of motion then contain terms of order $1/a$, where a is the radius of the (spherical) Earth. Moreover, if O is a point in North latitude ϕ taken at or near the 'centre' of the disturbance, S is the distance measured along a great circle on the Earth's surface from O to the foot of the perpendicular through a point Q in the atmosphere, θ is the angle between this great circle and the meridian of O, and Z is the perpendicular distance from Q to the Earth's surface, then the kinetic energy of unit mass of air is

$$K = \frac{1}{2}\left\{\dot{S}^2 + a^2\sin^2\frac{S}{a}\,\dot{\theta}^2 + \dot{Z}^2 + ka\sin\theta\,.\,\dot{S} + \left(\frac{ka^2}{2}\sin 2\frac{S}{a}\cos\theta + la^2\sin^2\frac{S}{a}\right)\dot{\theta}\right\},$$

where

$$k = 2\omega\cos\phi, \quad l = 2\omega\sin\phi,$$

and ω is the angular speed of the Earth about its polar axis. In this formula the small terms in ω^2 and Z/a are neglected. The corresponding kinetic metric is

$$ds^2 = dt^2 - \frac{1}{c^2}\left\{dS^2 + a^2\sin^2\frac{S}{a}\,d\theta^2 + dZ^2 + ka\sin\theta\,dS\,dt + \left(\frac{ka^2}{2}\sin 2\frac{S}{a}\cos\theta + la^2\sin^2\frac{S}{a}\right)d\theta\,dt\right\}. \tag{5·1}$$

Comparing (1·1) and (5·1) we make the identifications

$$x^1 = S, \quad x^2 = \theta, \quad x^3 = Z, \quad x^4 = t, \tag{5·2}$$

and

$$\left.\begin{gathered}\gamma_{44} = 0, \quad \gamma_{14} = \frac{ka}{2}\sin\theta, \quad \gamma_{24} = \frac{1}{2}\left(\frac{ka^2}{2}\sin 2\frac{S}{a}\cos\theta + la^2\sin^2\frac{S}{a}\right), \quad \gamma_{34} = 0, \\ \gamma_{11} = 1, \quad \gamma_{22} = a^2\sin^2\frac{S}{a}, \quad \gamma_{33} = 1, \quad \gamma_{pq} = 0 \quad \text{when} \quad p \neq q, \\ \sqrt{\Delta} = a\sin\frac{S}{a}.\end{gathered}\right\} \tag{5·3}$$

The co-ordinate-system is therefore spatially orthogonal and the equations of motion and of continuity are of the form (1·2) and (1·3). We take Σ_0 to be the Earth's surface and we observe that x^3 $(= Z)$ does not occur explicitly in the $\gamma_{\mu\nu}$; hence when we apply the transformation (2·4) to (5·1) we have, for the non-zero $a_{\mu\nu}$ in (2·5),

$$\left.\begin{aligned}a_{\mu\nu} &= a_{\mu\nu}(\alpha^1, \alpha^2), \\ a_{33} &= h_3^2 = 1.\end{aligned}\right\} \tag{5·4}$$

with

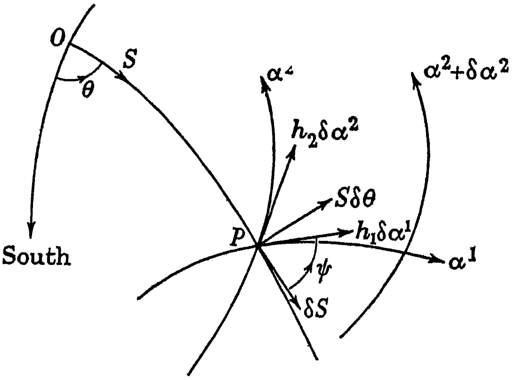


Figure 2

We take the case of cyclonic motion as standard and regard the α^2-curves as the isobars. Along the isobar through P (see figure 2), α^2 is a parameter increasing along the isobar $p =$ *constant* in the direction shown by the arrow. Similarly $\alpha^2 + \delta\alpha^2$ is a parameter increasing along the isobar $p + \delta p =$ *constant* and $\delta\alpha^2 > 0$, $\delta p > 0$. We define the gradient-wind at P as a steady wind blowing parallel to the Earth's surface for which, in (3·5) to (3·8),

$$V^1 = 0, \quad V^2 = G, \quad V^3 = 0, \quad \frac{dG}{dt} = 0, \tag{5·5}$$

G being positive in the direction of increasing α^2. Also from (2·4), (2·6) and (5·2) we obtain

$$a_{\mu\nu} = \gamma_{\lambda\tau}\frac{\partial x^\lambda}{\partial\alpha^\mu}\frac{\partial x^\tau}{\partial\alpha^\nu},$$

whence $a_{14} = \gamma_{14}\dfrac{\partial S}{\partial\alpha^1} + \gamma_{24}\dfrac{\partial\theta}{\partial\alpha^1}$, $\quad a_{24} = \gamma_{14}\dfrac{\partial S}{\partial\alpha^2} + \gamma_{24}\dfrac{\partial\theta}{\partial\alpha^2}$, $\quad a_{34} = 0$, $\quad a_{44} = 0$, (5·6)

since the functions χ_1, χ_2 of (2·4) do not contain α^3 $(= Z)$ explicitly.

We assume, as usual in gradient-wind theory, that gravity is the only external force per unit volume acting, so that

$$F_1 = 0, \quad F_2 = 0, \quad F_3 = -g,$$

and (3·7) then yields the hydrostatic equation

$$-\frac{1}{\rho}\frac{\partial p}{\partial Z} = g. \tag{5·7}$$

Since the α^2-curves are isobars, $\partial p/h_2\partial\alpha^2$ is zero and, by (5·4), (5·5) and (5·6), the equation (3·6) is an identity whilst (3·5) and (3·8) become, respectively,

$$\kappa_2 G^2 + \frac{G}{h_1 h_2}\left(\frac{\partial a_{24}}{\partial\alpha^1} - \frac{\partial a_{14}}{\partial\alpha^2}\right) = \frac{1}{\rho}\frac{\partial p}{h_1\partial\alpha^1}, \tag{5·8}$$

$$\frac{\partial\rho G}{h_2\partial\alpha^2} + \rho\kappa_1 G = 0, \tag{5·9}$$

where κ_2 is the geodesic curvature of the isobars and κ_1, that of the orthogonal trajectories to the isobars. But now since G is not an explicit function of t, (5·5) yields

$$\frac{dG}{dt} = G\frac{\partial G}{h_2\partial\alpha^2} = 0,$$

and G is therefore constant along an isobar. The last equation together with (5·9) gives

$$\frac{\partial\log\rho}{h_2\partial\alpha^2} + \kappa_1 = 0.$$

If at this point we were to assume, as is so often done in dynamical meteorology, that the horizontal variations of the air density were negligible, we should have $\kappa_1 = 0$ which would mean that the curves orthogonal to the isobars *were always great circles*. In practice, there is no reason for supposing that this restriction holds good and consequently, the existence of a gradient wind implies in general that horizontal variations of air density are appreciable.

In (5·8), $\partial p/h_1\partial\alpha^1$ is the pressure-gradient normal to the isobar and it is positive towards the region of *high* pressure. The coefficient of G on the left-hand side is transformed as follows by using (5·6) and (5·3):

$$\begin{aligned}\frac{1}{h_1 h_2}\left(\frac{\partial a_{24}}{\partial\alpha^1} - \frac{\partial a_{14}}{\partial\alpha^2}\right) &= \left(\frac{\partial\gamma_{24}}{\partial S} - \frac{\partial\gamma_{14}}{\partial\theta}\right)\left(\frac{\partial S}{\partial\alpha^1}\frac{\partial\theta}{\partial\alpha^2} - \frac{\partial S}{\partial\alpha^2}\frac{\partial\theta}{\partial\alpha^1}\right)\frac{1}{h_1 h_2} \\ &= \left(\frac{ka}{2}\cos 2\frac{S}{a}\cos\theta + la\sin\frac{S}{a}\cos\frac{S}{a} - \frac{ka}{2}\cos\theta\right)\left(\frac{\partial S}{h_1\partial\alpha^1}\frac{\partial\theta}{h_2\partial\alpha^2} - \frac{\partial S}{h_2\partial\alpha^2}\frac{\partial\theta}{h_1\partial\alpha^1}\right).\end{aligned} \tag{5·10}$$

But now referring to figure 2 we observe that

$$\frac{\partial S}{h_1\partial\alpha^1}\frac{S\partial\theta}{h_2\partial\alpha^2} - \frac{\partial S}{h_2\partial\alpha^2}\frac{S\partial\theta}{h_1\partial\alpha^1} = \cos\psi\cos\psi - (-\sin\psi)\sin\psi = 1. \tag{5·11}$$

Moreover if in (5·10) we neglect the terms of order $1/a^2$ and higher, we obtain by (5·11)

$$\frac{1}{h_1 h_2}\left(\frac{\partial a_{24}}{\partial \alpha^1}-\frac{\partial a_{14}}{\partial \alpha^2}\right) = l-\frac{kS}{a}\cos\theta.$$

Hence our equation (5·8) for a large-scale gradient wind becomes

$$\kappa_2 G^2+G\left(l-\frac{kS}{a}\cos\theta\right) = \frac{1}{\rho}\frac{\partial p}{h_1\partial\alpha^1},$$

which replaces the equation $\kappa_2 G^2+lG = \rho^{-1}\partial p/h_1\partial\alpha^1$, usually given in the text-books which is valid (McVittie 1948) only for small-scale circular motion about O. We observe that the curvatures κ_1 and κ_2 are not independent but by (3·9) satisfy

$$\frac{\partial\kappa_2}{h_1\partial\alpha^1}+\frac{\partial\kappa_1}{h_2\partial\alpha^2}+(\kappa_1)^2+(\kappa_2)^2 = -\frac{1}{a^2},$$

since, for a sphere of radius a, $K_0 = a^{-2}$.

We shall not attempt to discuss here the thermal history of unit mass of air in a motion of this kind beyond saying that a gradient wind on a large scale is impossible unless there is some mechanism such as conduction for adjusting the heat-content of the unit mass during its motion (McVittie 1948).

REFERENCES

Eisenhart, L. P. 1940 *Introduction to differential geometry*, §§ 28, 34. Princeton Univ. Press.

Meyer, R. E. 1948 *Quart. J. Mech. App. Math.* **1**, p. 196.

McConnell, A. J. 1931 *Applications of the absolute differential calculus*, pp. 143–145, 169, 184. London and Glasgow: Blackie and Son, Ltd.

McVittie, G. C. 1948 *Quart. J. Mech. App. Math.* **1**, 174, §§ 7, 8.

McVittie, G. C. 1949 *Proc. Roy. Soc.* A, **196**, 285.

BAKERIAN LECTURE

Nuclear explosions

By Sir George Thomson, F.R.S.
Imperial College of Science and Technology, London

(*Delivered* 24 *June* 1948—*Received* 24 *August* 1948)

[Plates 4 to 7]

FOREWORD

There is an old Scotch proverb 'Fools and bairns should not see a thing half done'. An audience of Fellows of the Royal Society certainly comes into neither category, so I make no apology for presenting to you to-day a subject which is probably nearer its beginning than its end, a subject in which most of the conclusions are tentative. In the course of what I have to say I shall ask many more questions than I can answer, and even such answers as I can suggest will be somewhat speculative.

INTRODUCTION

The work I have to describe to-day comes from a discovery by two Viennese ladies, Blau & Wambacher (1937). They found that photographic plates kept for some time showed 'stars', i.e. groups of tracks recalling those in a Wilson chamber, but reduced in proportion to the greater density of the photographic plate compared with the gas in the chamber. The tracks in a star diverge from a point in the emulsion. Some of the stars were due to contamination and represented successive alpha-particle disintegrations due to a speck of radioactive impurity in the plate. But others included tracks too long to be due to any known radioactive impurities. Furthermore, Stetter & Wambacher (1938, 1939) showed that the number of stars increased if the plates were taken for a few weeks to the top of a mountain. This was correctly interpreted as meaning that they are due to some component of the cosmic rays and represent fragments from the exploding nucleus of some atom in the emulsion. Since then many investigations have been made on these lines, especially by the group under Professor Powell at Bristol (Powell & Occhialini 1947*a*). The technique owes much to the admirable work of the research staff of Ilfords and more recently that of Kodaks. To-day I want to discuss the nature and origin of these stars and their relation to analogous phenomena observed in the Wilson chamber. I shall illustrate the arguments from data obtained at Imperial College by Messrs Perkins, Lattimore, Harding, Li and others, to whom I am much indebted for permission to use some still unpublished work.

Stars contain anything up to thirty-four tracks; one of twenty-five has been observed at Imperial College. The tracks vary in length, in specific ionization and in straightness. It is possible in favourable cases to distinguish between tracks due to alpha-particles, protons, mesons and heavily charged particles. Deuterons and tritons also occur. The plates are not sensitive to particles with low specific ionization. In practice this means all electrons (though Kodaks have recently produced a plate sensitive to electrons of not more than a few kilovolts energy), mesons over about 8 MeV, protons over 100 MeV and alphas over several thousand MeV. These figures apply to particles from known stars which can be examined carefully; a large proportion of single tracks of these densities would probably be missed. Neutral particles are of course only detectable by knock-ons or by secondary disintegrations which will not normally occur near the point of origin, so they cannot be ascribed to any particular star.

EXPERIMENTAL DATA

Data on stars have been collected by Blau & Wambacher (1937) and other workers. I propose to show in most cases results from plates examined at Imperial College, partly because with improved technique they are likely to be more accurate than much of the earlier work, and partly because it is an advantage to have all the data derived from plates treated and examined in the same way.

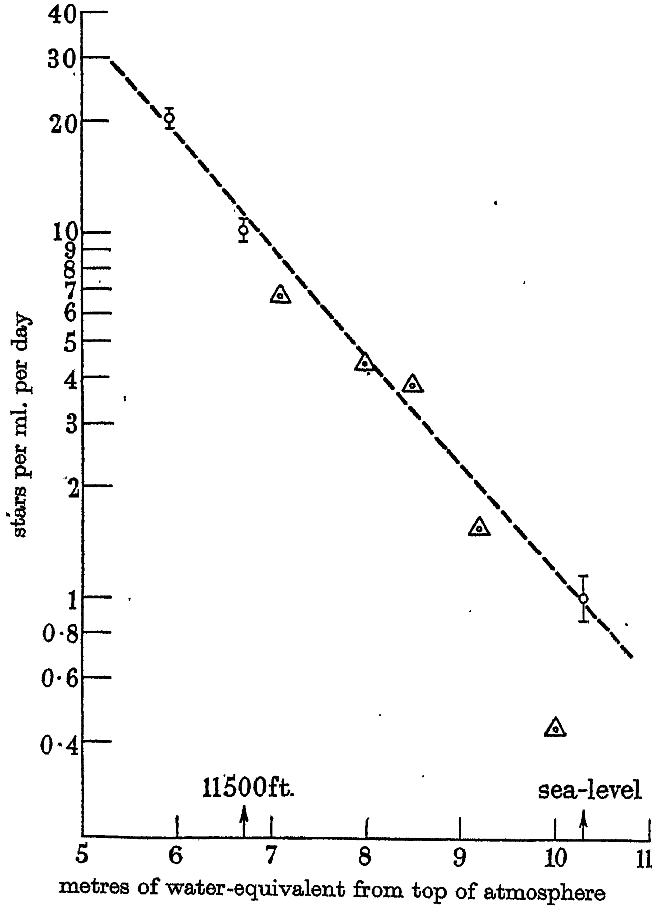


FIGURE 1. Altitude increase of stars. △ Stetter & Wambacher (normalized at 11,500 ft.). ⚲ Perkins.

Figure 1 shows the way in which a number of stars produced in a given time in a plate of given thickness varies with the height. The results of Perkins and of Powell (not shown in figure 1) agree well at 11,000 ft. The results of Stetter & Wambacher are adjusted to fit at this altitude. It will be seen that down to sea-level there is an approximately logarithmic variation corresponding to an exponential of seven for the whole of the atmosphere. This is about the same as for the soft component of the cosmic rays, and also for the neutrons which have been observed in them.

Stetter & Wambacher's results at sea-level are probably low relative to those at greater altitude because of fading of the image.

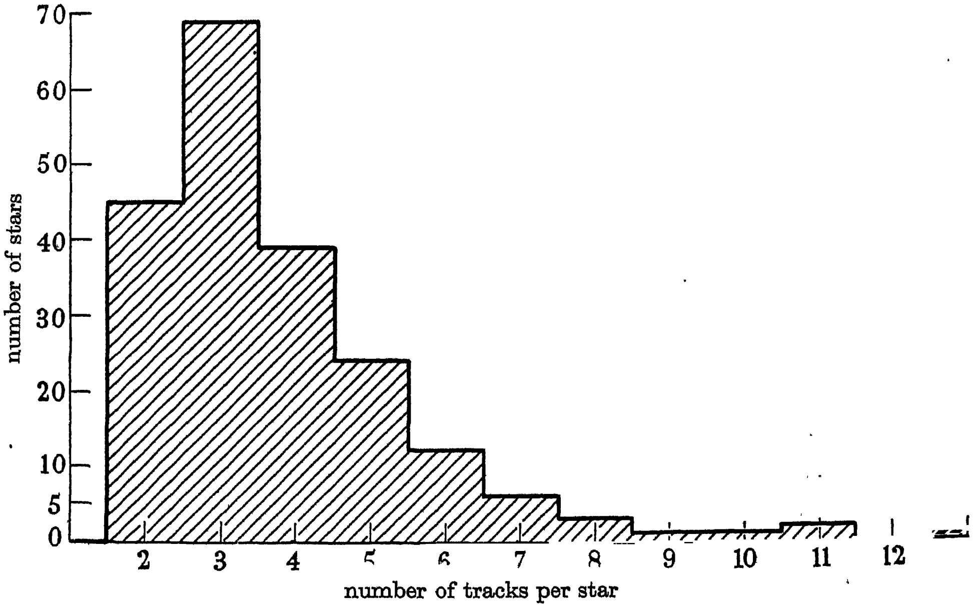


Figure 2. Frequency distribution of tracks from stars (3600 m.).

Figure 2 shows the frequency distribution of tracks among the stars. It refers to stars found on plates exposed at 3500 m. near Chamonix. It will be seen that the greatest number of the stars have three tracks and that few have more than about ten. Fewer stars are recorded with two than with three tracks. It is possible that the number of stars with two tracks is considerably underestimated. Such stars cannot in practice be distinguished from a single track which has been scattered at one point through a large angle. Further, those stars whose tracks, when projected on the photographic plate, are nearly in the same direction will not be counted.

Figure 3 shows the total energy associated with the charged particles from a group of fifty stars measured by Mr Li and Mr Harding. Binding energy of a proton has been taken as 8 MeV, and the energy of the protons has been doubled to allow for that of the neutrons which presumably accompanied the protons. These stars were selected at random and their number of tracks varied from two to five.

Though the large stars with more than ten tracks are few in number they are especially interesting, and Mr Perkins has made a detailed study of them. Figure 1a* shows an energy distribution of the protons for all the stars with over fifteen tracks which have been found so far at Imperial College, and figure 1b* shows a corresponding diagram for those with between ten and fourteen tracks, obtained by Messrs Harding and Lattimore.

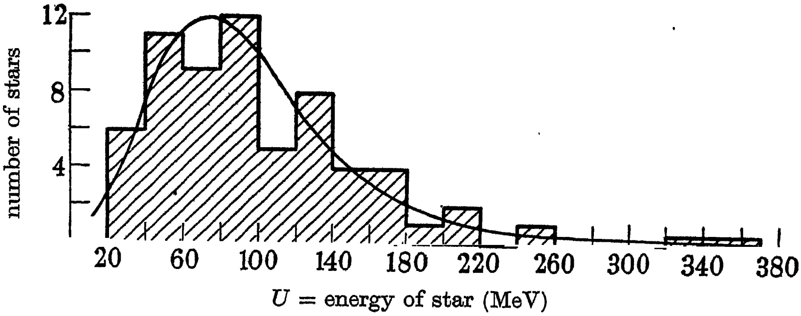


FIGURE 3

The distribution of energy can be partially explained on the supposition that the nucleus is struck by a particle which communicates its energy to the individual nucleons so that a state is set up which can be regarded as one of high temperature. Nucleons will escape in a process akin to evaporation. As this process proceeds both temperature and effective atomic weight of the nucleus will diminish. Classical theory (Weisskopf 1937) indicates a relation between the temperature and the energy spectrum of the protons as follows:

$$P(E)\,dE = \frac{E-V}{T^2}\,e^{-(E-V)/T}\,dE,$$

where E is the energy, V the potential barrier and T the temperature, all in MeV. This is somewhat modified by quantum considerations concerning the penetrability of the barrier by particles with different angular momentum.

The relation between the temperature and energy of excitation depends on what model is assumed for the nucleus, but the differences are fortunately not very large. Calculations on these lines were first carried out by Bagge (1941), but there appears to have been an error in his method of normalization. Perkins's calculations show a reasonable agreement with theory, over the greater part of the range of energy, assuming that the relation between temperature and excitation of the nucleus is that for a Fermi gas, namely, $U \propto T^2$. There is, however, a considerable excess of protons of very high energy. Mr Perkins considers that these are produced by a process suggested by Heisenberg (1938), in which the particle strikes off a few of the nucleons nearest its point of contact by a glancing blow, so that there is no time for them to get into thermal equilibrium with the rest of the nucleus. Besides the protons, and a few deuterons and tritons not easily distinguishable from them, there are present in the stars about 30 % of alpha-ray tracks. The spectrum of these tracks is shown in figure 2*. These tracks show little variation in grain density, and it is not therefore possible to estimate the energy of the particle producing a track, unless the track ends in the emulsion. Since, of course, a particle with low energy is more likely to do so than one with large energy the spectrum is distorted, for only those tracks which stop in the emulsion can have their energy measured. To allow for this the calculated curves are corrected by a 'loss factor' as shown in the diagram. The alpha-particles also show a proportion of fast particles.

---

* Figure numbers marked with an asterisk refer to figures in the paper by Harding *et al.* which follows this paper on p. 325.

It will be noticed that the 'temperature' which gives the best fit for the alphas is 7 MeV, that for the protons is 5 MeV for the stars of ten to fourteen particles and 8 MeV for those over fifteen. This is a tolerable agreement, since the curve for the alphas is the mean for tracks for all the stars. On the other hand, the value of V is quite different, being 5 MeV for the protons and only about 3 MeV for the alphas, while owing to the double charge on the latter it should be twice as large as for the protons. This is an indication that the alphas are emitted from a portion only of the nucleus with a lower potential barrier.

Figure 1 *d**, for which I am indebted to Mr Li, shows the energy distribution from the smaller stars. It also shows agreement with an evaporation formula except for 'a tail' of high-energy particles which are proportionally less numerous than in the case of the large stars. The temperature also is lower than for the larger stars. This seems to be part of a general phenomenon by which the larger the star the higher the effective temperature, and is what one would expect. The large stars must have had a large excitation energy and this implies a high temperature.

The angular distribution of the particles of the large stars shows some very interesting features. Figures 5*, 9* and 10* show the results of an analysis made as follows:

Each track is treated as a vector of unit length, irrespective of the energy of the track, and the resultant of these vectors found for each star, taking separately the proton and alpha tracks. Assuming that the tracks were really directed at random, it is possible to calculate the distribution of these results. We have, in fact, an example of the well-known 'random-walk' class of problem. It will be seen from the diagrams that while the distribution of the protons is approximately what would be expected, and that the same is true for the alpha-particles from the stars of ten to fourteen tracks, the alpha-particles from the large stars show a very unexpected asymmetrical distribution. That this asymmetry is a real one can be seen by considering the relation between the direction of the resultant vector, and that of the heavy track which represents the remainder of the nucleus. In figures 10 *a** and 10 *b** are histograms showing the number of stars for which the angle between these two vectors has particular values. On random distribution we should expect a distribution similar to the sine curve which is drawn in. It will be seen that the alpha-particles from the large stars again show an asymmetrical distribution confirming the results of the previous investigation. To explain this, Perkins has put forward the view that the nucleus forming these large stars splits into two, and that the portion which gives rise to the alpha-particles is the smaller fragment travelling with considerable speed with regard to the common centre of gravity which is nearly at rest with respect to the rest of the emulsion. In this connexion it is worth noticing that some large stars show *two* short heavy tracks, corresponding presumably to the two parts of a residue which has divided.

---

* Figure numbers marked with an asterisk refer to figures in the paper by Harding *et al.* which follows this paper on p. 325.

SINGLE TRACKS

A considerable number of single tracks accompany the stars on the plate. Perkins has shown that this ratio is constant from sea-level to 4,300 m. It is not appreciably changed if a layer of graphite, paraffin wax, brass or lead is placed in contact with the plate. Of the tracks about 2 % are due to alpha-particles, the remainder appear to be protons, though it is not always possible to distinguish these from deuterons or mesons. Of the tracks from stars 30 % are alphas, but these have only about one-quarter the range of the protons.

These facts are compatible with the view that the single tracks are parts of stars occurring outside the plates, and Perkins's earlier measurements seemed to show that the number of tracks relative to the number of stars was about that to be expected on this view. Re-examination of the plates with improved technique has shown that the number of tracks was considerably underestimated though the relative numbers were correct. While there is some uncertainty as to the number of high-speed protons in the stars, the most reasonable estimate only accounts for about twenty of the 110 tracks found per star in 100μ plates.

It is a matter of interest to consider to what the remainder are due. The obvious source is the neutrons, which, though invisible, are presumably emitted in the stars to keep normal the balance between charge and mass in the residue. These neutrons will produce knock-ons in the emulsion and in any material containing hydrogen outside it.

The neutrons will lose energy by inelastic collisions with the nitrogen in the air, and to a less extent by the disintegrations they cause in it. I will take $1{\cdot}2 \times 10^{-24}$ as the cross-section for a collision which prevents the neutron from producing a visible knock-on in the emulsion. This corresponds to a free path of 300 m. in air at sea-level. The cross-section for collision with a proton by 10 MeV neutrons is $0{\cdot}8 \times 10^{-24}$, and the number of protons per cm.^{3} of emulsion is 3×10^{22}, so the free path for producing knock-ons in the emulsion is $1/(0{\cdot}8 \times 10^{-24} \times 3 \times 10^{22}) = 40$ cm.

Suppose that the production of stars is proportional to the density of the material divided by the cube root of atomic weight and is $N\rho A^{-\frac{1}{3}}$. Let the number of neutrons per star be ν. Then the number of neutrons crossing the plate per sq.cm.sec. = $\nu\rho_{\text{air}}\dfrac{N}{14^{\frac{1}{3}}}\dfrac{\lambda'}{p}$, where p is a factor to allow for obliquity and $\lambda' = 3 \times 10^4$ cm. is the free path in air. The number of knock-ons is $\nu\rho_{\text{air}}\dfrac{N}{14^{\frac{1}{3}}}\dfrac{\lambda'}{p}\dfrac{pd}{\lambda}$, where d is the thickness of emulsion and λ the free path for knock-ons.

The number of stars per sq.cm. per sec. is $\rho_{\text{em.}}Nd/62^{\frac{1}{3}}$, 62 being the (weighted) mean atomic weight for the emulsion. Hence the ratio

$$\frac{\text{knock-ons}}{\text{stars}} = \nu\left(\frac{62}{14}\right)^{\frac{1}{3}}\frac{\rho_{\text{air}}}{\rho_{\text{em.}}}\frac{\lambda'}{\lambda} = \nu\frac{1{\cdot}3 \times 10^{-3}}{4}\frac{3 \times 10^4}{40}1{\cdot}62 \doteqdot \nu/2{\cdot}5.$$

We should not expect ν to exceed about 4, but measurements have been made by Korff (1939) and by Halban, Kowarski & Magat (1939) of the number of neutrons in the cosmic rays. A detailed analysis by Bethe, Korff & Placzek (1940) leads to the conclusion that the production of neutrons is 0·05/g./sec. at a height where the density is 0·1 atm., and that it varies with height with an exponential of 7 per atm., which is nearly the same as for stars. At sea-level this would be about 5×10^{-5}/g./sec. in air, as against about $2·5 \times 10^{-6}$ stars/g./sec. in emulsion. Allowing for the difference in atomic weight, this gives fourteen neutrons per star, which may either indicate that there is another source of neutrons or that the number of very small stars has been much underestimated.* Independent neutrons of course would add to the knock-on effect unless they are released with rather low energies.

Even if we suppose that all the neutrons observed by Korff and others have an energy of 10 MeV, they would not account for more than about 4% of the observed tracks. Korff detected his neutrons when they had been reduced to relatively small energies. If any neutrons started with very high energies they might each produce a number of knock-ons before being slowed down, or they might produce some type of disintegration (perhaps even stars) before reaching the energies at which Korff would have detected them.

However, fast neutrons, or indeed any neutrons, as a major cause of the single tracks seem excluded by the experiments of Perkins, which show no difference in the number of tracks between plates surrounded by lead and by paraffin, unless, indeed, we suppose that the protons start with so much energy that the material in the immediate neighbourhood furnishes only a small proportion of those observed in the plates.

If the observed protons start with energies corresponding to more than about 1 cm. range in paraffin they would escape detection in the emulsion till partly slowed down. The number of such particles with a given residual range should be independent of the range. Grain counts on the single tracks seem to show a preference for energies of a few million volts, but this work is still in progress.

Accurate knowledge of the range distribution would also test the hypothesis that the extra single tracks are due to fast particles from stars which escape observation owing to their speed, or which is practically equivalent, as far as this phenomenon is concerned, to single protons released with high energy from nuclei by some unknown cause which might or might not be that which forms the visible stars. Since only about 4 % of the tracks *start* in the emulsion we can exclude 'single-track' stars with a normal energy distribution.

It is worth seeing if these single tracks could be due to protons coming down from the top of the atmosphere, but the following considerations tell against this view. The number of proton tracks stopping in the plate is about five times that of the meson tracks which do so. Hence for equal absorption of protons and mesons fivesixths of the penetrating component of the cosmic rays would have to be protons. It is true that the absorption of the penetrating component is less than that of the radiation responsible for the proton tracks as judged from the way in which the two vary in intensity with altitude, but the factor is only of the order of 2, and this still leaves far too many protons to be reconciled with Wilson chamber experiments.

---

* *Note added in proof*, 17 *November* 1948. Recent measurements by Yuan (1948) confirm earlier results by Korff & Hamermesh (1946) and Agnew, Bright & Froman (1947) to give a much lower value for the production of neutrons. Extrapolated to sea-level their results give about 1×10^{-5}/g./sec. in air. Allowing for the difference in atomic weight this gives about $1\frac{1}{2}$ neutrons per star, which is a rather low number. The above argument as to the insufficiency of neutrons to cause the single tracks holds with the new data *a fortiori*.

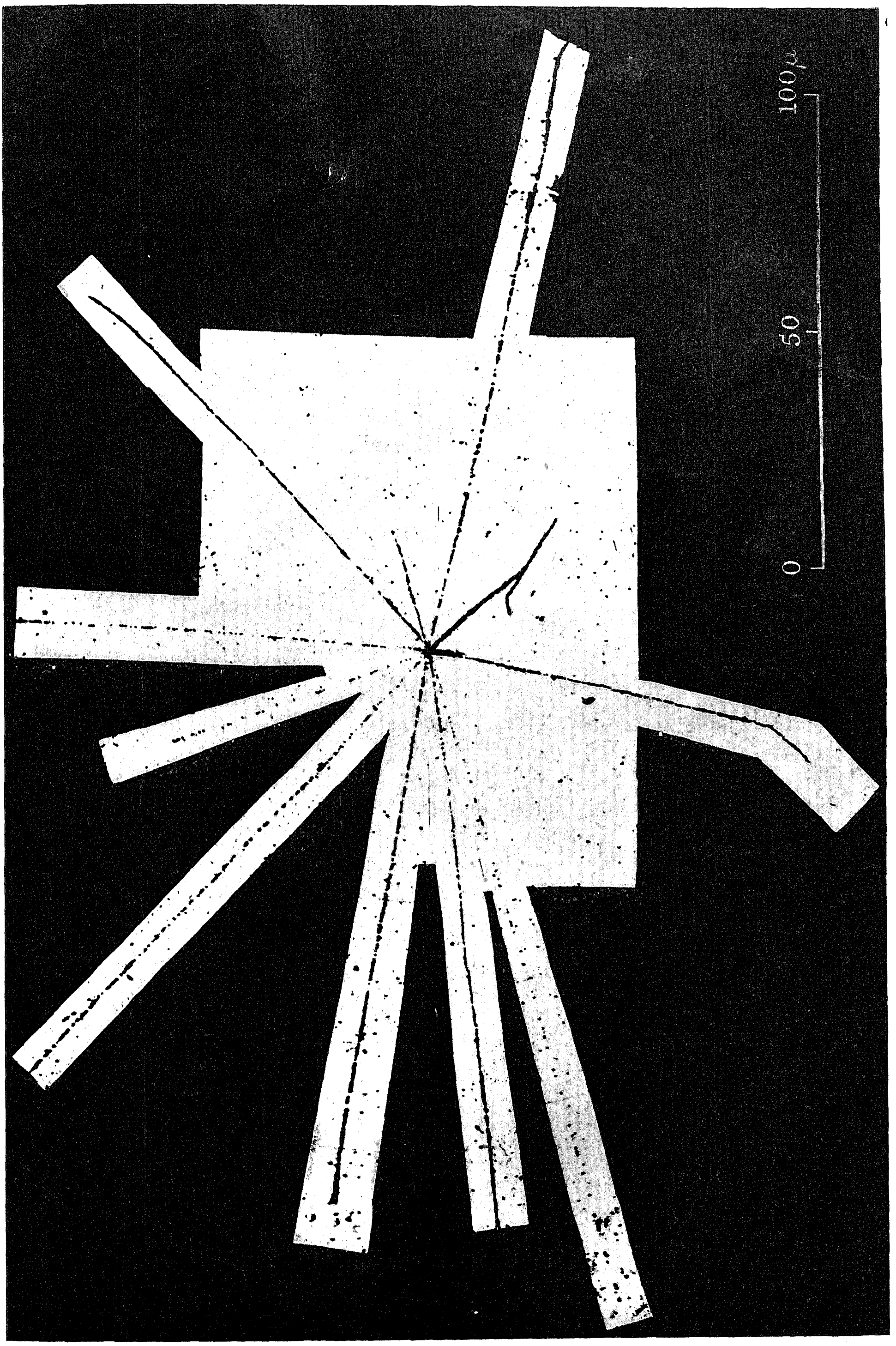
Finally, it may be asked, can we be sure that a large number of these single tracks are not due to mesons? It can be shown that, if n particles are being brought to rest per cm.^{3} per sec. in a medium, and if these particles leave visible tracks for the last L cm. of their course, if they are isotropically distributed in angle, and if the relevant properties of the medium are the same throughout, then the number of tracks per sq.cm. in a thin layer in the medium is $\frac{1}{2}nL$. Now the number of mesons stopped per cm.^{3} per day is of the order of unity at sea-level (as deduced from cosmic-ray data), and L for mesons is about $800\mu = 0{\cdot}08$ cm., which gives 0·04/sq.cm./day for the single tracks, while the observed value is about unity. The same formula, used to compare the number of single tracks with those seen to stop in the emulsion, gives $L = 0{\cdot}4$ cm., which is about right for protons and verifies the conclusion that the single tracks are indeed mostly these. It seems that for the present we must suspend judgement as to their origin.

WILSON CHAMBER WORK

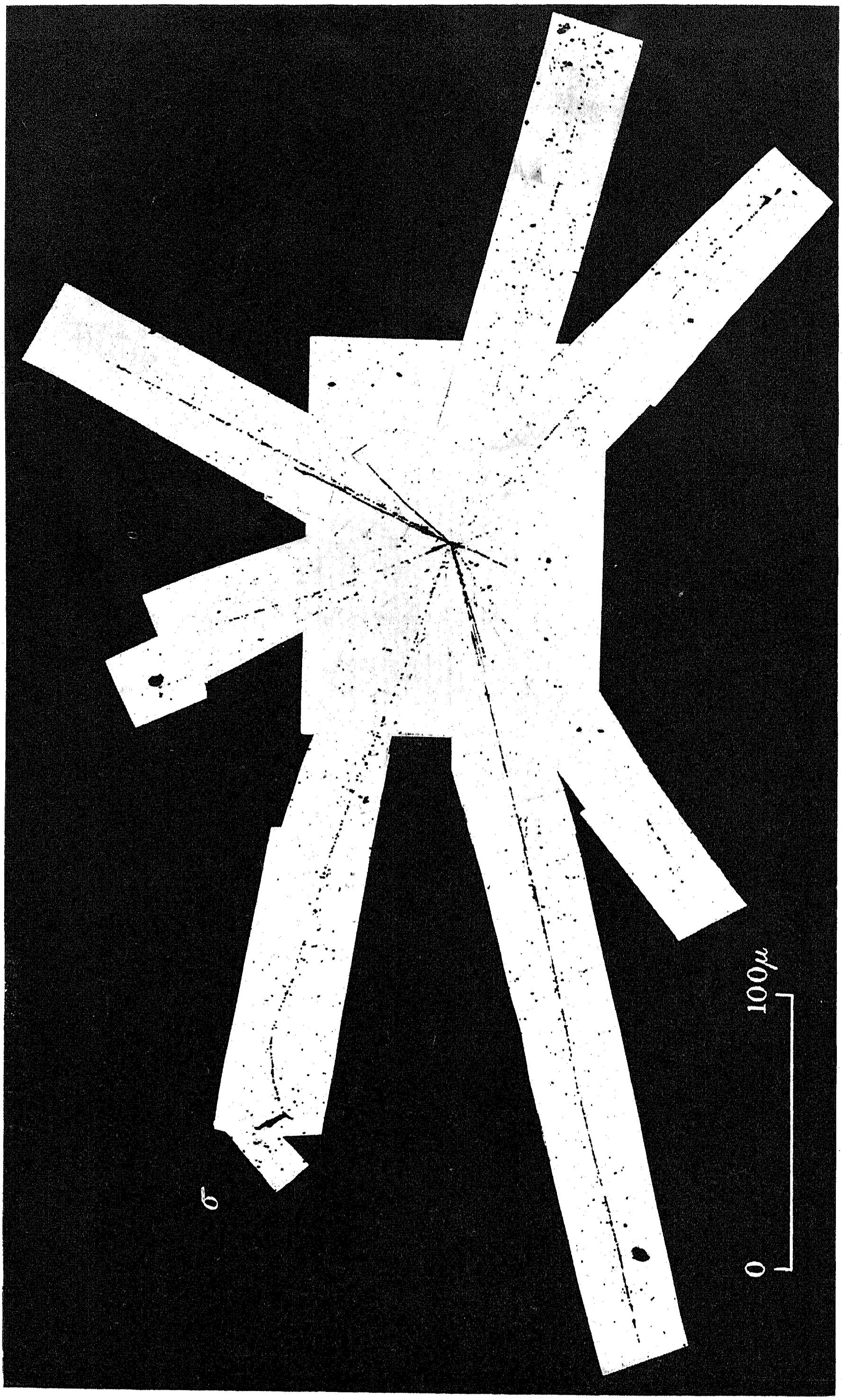
Effects have been observed in Wilson chambers which are undoubtedly closely connected with the stars seen in photographic plates. They have been specially studied by Hazen (1944) and by W. M. Powell in U.S.A. (1946). Both these workers used a Wilson chamber crossed horizontally by a number of lead plates of 0·7 and 1·0 cm. thickness respectively. They worked at altitudes of 10,000 and 13,400 ft. In each case they found a few stars in the gas, many more coming out of the lead. Their stars often contained electrons, and the majority of them had most or all of their tracks directed downwards.

In comparing these results with those in photographic plates it is important to consider the differences in the conditions under which the tracks in the stars became visible in two cases. Thus the photographic plate does not show electrons at all, nor heavy particles of more than a certain energy. For a proton this is roughly what is required to penetrate one of the lead plates used in the Wilson chamber experiments. A fair proportion of the stars observed had some or all of their particles capable of doing this, and some particles were able to pierce three or four plates. Since the mass in the plates far exceeds that in the gas one would expect most of the stars to be formed in the lead, and this agrees with what is observed, but the actual discrepancy is undoubtedly even greater, since many stars formed in the lead will be wholly absorbed in it and never seen at all. Thus the photographic technique selects the stars with slow particles, while the Wilson chamber selects those whose particles are fast.

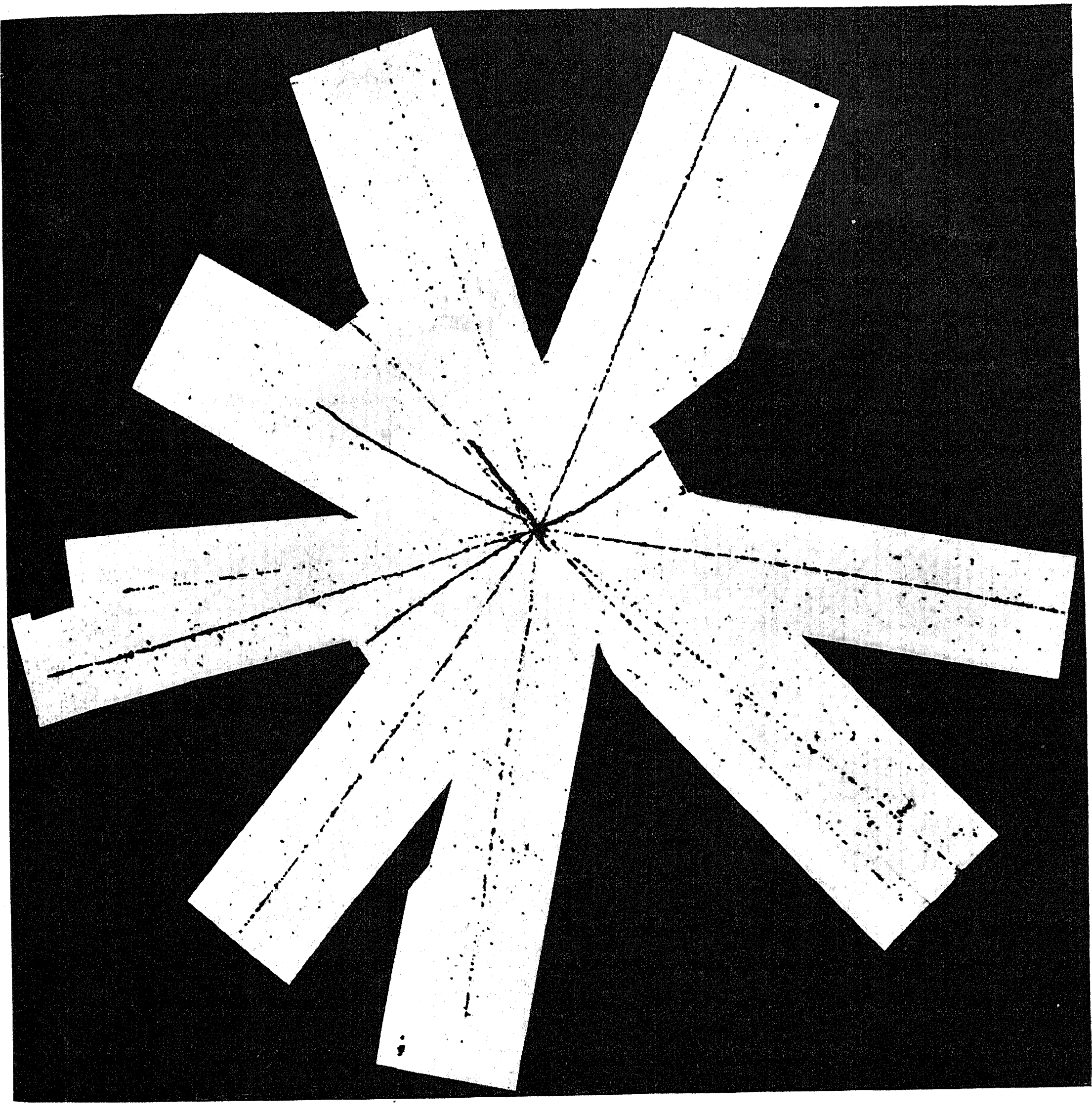
The Wilson chamber experiments agree in showing an average of about one electron or meson track per star. This greatly exceeds the small proportion of mesons



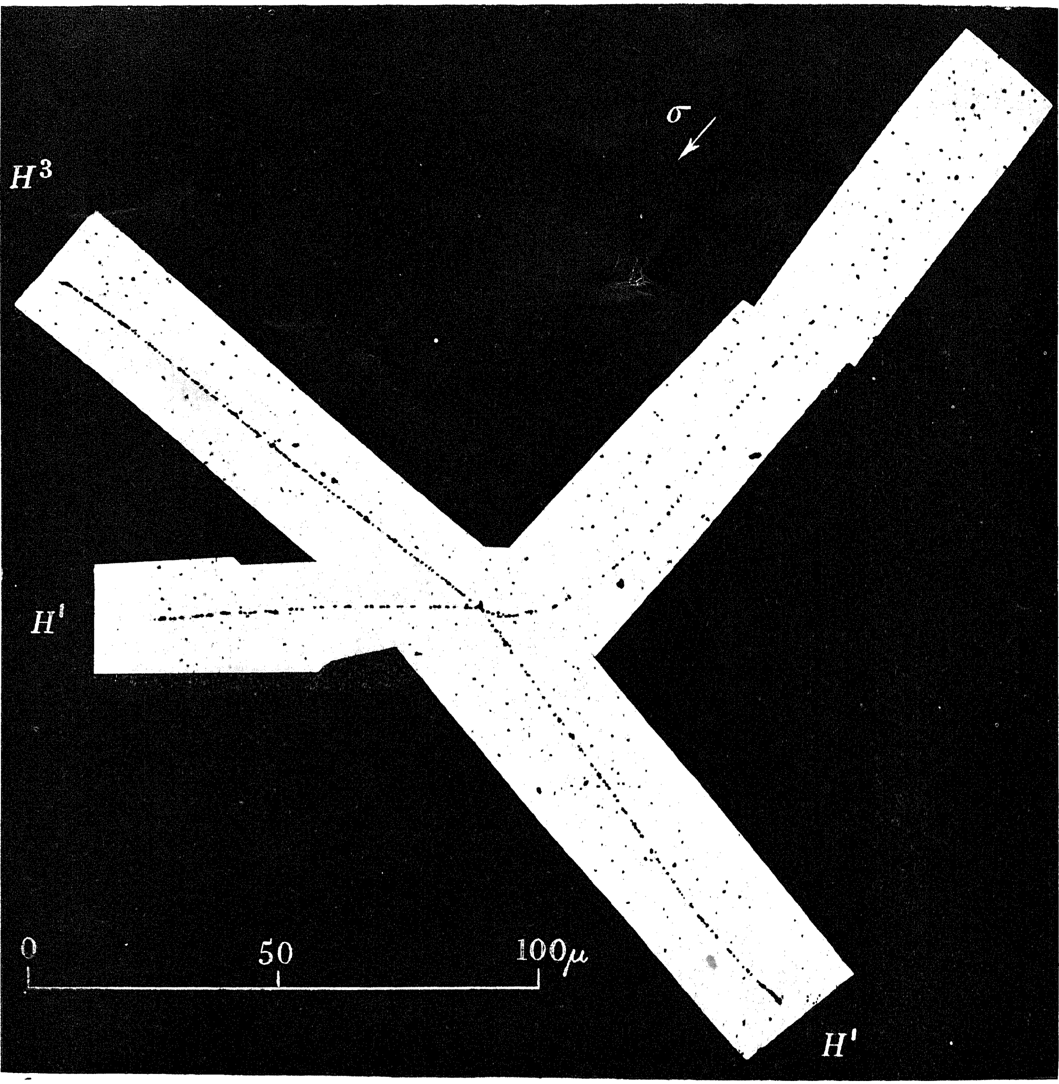
Photograph of a disintegration leading to the ejection of protons and alpha-particles, and of a nucleus of Li_3^8, which comes to rest in the emulsion and subsequently undergoes β-decay. The resulting Be_4^8 nucleus disintegrates spontaneously into two alpha-particles moving in opposite directions, giving the appearance of a T-track. 100μ C2 emulsion.



An example of a distintegration giving rise to fourteen tracks, one of which (σ) has the characteristics of a slow meson. This ejected meson terminates its range in the emulsion layer and undergoes nuclear capture, with the production of a second star consisting of three very short tracks. 100μ C2 emulsion, boron-loaded.



Photomicrograph of typical star resulting from the energetic distintegration of a silver or bromine nucleus. All or most of the lightly ionized tracks are due to protons, and the very dense tracks to alpha-particles or heavier nuclei. The very short, slightly curved track is attributed to the recoil fragment. 100μ C2 emulsion, boron-loaded.



Star resulting from nuclear capture of a σ meson. Three charged secondaries are produced. Recorded in 40μ B1 emulsion exposed in high-altitude aircraft.

seen in stars in photographic plates; the balance may represent particles which ionize too little to be visible there, or may be due to the Wilson chamber stars being a different sample from those in the photographic plates.

Few of the stars in the plates show strong downward concentration of the tracks. In contrast, this is the rule in the chamber stars. Thus Hazen (1944), who divides his stars into two roughly equal groups according as to whether any of the tracks in them can penetrate 7 mm. of lead or not, finds that in the penetrating class only six out of twenty-seven have any upward particles at all, while even in the non-penetrating class only eleven out of thirty-one had about equal numbers of up and down tracks.

It is possible to estimate from Hazen's data the proportion which the penetrating class bear to the whole, including those that do not emerge at all, if we suppose that stars are seen from a depth of 5×10^{-3} cm. in the lead, which is about what one would expect if an average star is composed of three or four particles each capable of going 0·03 cm. through lead, the range of a proton of 10 MeV. It comes to 3·3 %. Probably many of these stars are wholly composed of long-range particles and would be missed altogether in the plates, so their absence is not disturbing. It does, however, seem odd that the large proportion of the less penetrating stars which have their tracks strongly directed downwards should be so little represented on the plates. There seems to be a continuous gradation in the stars from symmetrical stars of low average track length to downward-directed stars with long tracks. No doubt therefore the downward-directed stars even in Hazen's less penetrating group have longer tracks than the average. They can thus more easily escape from the lead plate and are over-represented. Even if one makes the extreme and obviously untrue assumption that all such stars are actually counted the ratio is 2·5 %, which should be observed. It should, however, in fairness be said that Hazen, by comparing the ratio of the number of stars observed in air and in the lead with the assumed stopping power gets a much larger number for the stars which escape detection. As, however, he only observed two stars in the air the statistical uncertainty of this method of calculation is large.

From Powell's results a similar calculation can be made, though he divides his stars in a different way. He records sixteen 'upward' stars out of a total of 156. On the same assumptions as to range, etc., the proportion of non-isotropic (supposed all counted) to isotropic is 4 %. In Powell's case it is possible to check by calculating the number of stars in the gas and comparing with the observed number, namely, thirteen, which is large enough to give reasonable statistics. The result comes too small by a factor of 2·5, assuming the relative cross-section for star production in lead and argon to be in the ratio of the areas of their nuclei. The most likely explanation is that some of the smaller stars coming from the lead are missed, especially those which have only a few particles directed outwards. If this is the case the true proportion of non-isotropic stars will be less than we have calculated above.

Assuming that the chance of star production in an atom of lead is to that in an atom of argon, the gas used in the chamber, as the areas of the nuclei, we can calculate from the observed number of stars in the gas the total number which must have been formed in the lead. Nearly all the penetrating ones will have been observed and, assuming that they have, we find that they comprise 1·5 % of the whole. Perhaps half of these will consist of particles of more energy than can be detected in a photographic plate, leaving a residuum which would require rather refined statistics to detect, since of course some stars are bound to show collimation in a downward direction as a mere result of a random distribution.

FREQUENCY OF STAR PRODUCTION IN DIFFERENT MEDIA

It is a matter of importance to determine how the chance of a star being formed depends on the character of the nucleus that may form it.

We now have plates made by Ilfords in which layers of gelatine free from silver bromide are interleaved with layers of emulsion, and others with layers containing lead. From observations of the stars in these plates, now in progress at Imperial College, it should be possible to determine with greater accuracy and certainty how this important quantity varies with the atomic weight.

Such evidence as exists at present is compatible with the view that the cross-section for star production varies as the cross-section of the atomic nucleus, i.e. about as the $\frac{2}{3}$ power of the atomic weight.

ARTIFICIAL PRODUCTION OF STARS

Recent work at Berkeley, California, with the new giant cyclotron has shown that effects closely resembling stars can be produced artificially, and it has also been found possible to create mesons of at least two masses. This work has not yet been reported in detail so it is impossible to discuss it fully.

The stars were of course all produced by ionizing radiation (deuterons and alpha-particles) and they seem to resemble more closely the collimated stars of the Wilson chamber, many of which are produced by ionizing radiation, than the more isotropic stars of the photographic plates. In particular, it is recorded (Gardner & Paterson 1948) that in the stars initiated by alpha-particles about three times as many tracks occur in a forward as in a backward direction. About three tracks per star were found on an average, but the plates were not sensitive to protons with more than about 10 MeV.

MESONS IN STARS

Some stars are caused by mesons. The first example of this was found by Perkins, and a number of others have been found since, especially by Powell, who calls the mesons which cause disintegrations of this type σ mesons (Powell & Occhialini 1947*b*). Powell has also shown that mesons can be produced by stars and that some of these are capable of producing stars themselves, i.e. are σ mesons (Lattes, Occhialini & Powell 1947).

It is not an easy matter to distinguish between the tracks due to protons and to mesons. The scattering is greater for mesons of the same velocity, and if the track ends in the emulsion the difference between the grain density at equal distances from the end is marked. It is, however, quite possible that some of the tracks which are attributed to fast protons leaving the emulsion are really due to mesons. For this reason no definite figure can be given for the number of mesons produced per star. A *lower limit* is 1 %.

About one star in thirty is due to a σ meson. Some fifteen well-marked examples have been studied by Perkins and Lattimore in an attempt to find out the mass of the meson concerned (Lattimore 1948). Values of 270 ± 70 and 290 ± 80 have been found by different methods for the *mean* of these. The individual spread is considerable, and it is not possible to say whether all these mesons have the same mass, but if they have it must be more than 200, which is the mass of the common mesons in the penetrating component.

THE ORIGIN OF THE STARS

The origin of these stars is a fascinating problem, and one which is still open to many speculations. There are, however, certain facts which enable us to rule out some of the causes which have been suggested. In the first place, the work with Wilson chambers shows that the stars are not correlated with cascade showers, not even with the very large Auger showers. It is fair to conclude from this that the stars are not due to photons or electrons. The Wilson chamber work also shows that few of the isotropic stars are due to ionizing radiation of any kind. We have seen that about one star in thirty is due to mesons, and this would roughly account for the proportion seen in the Wilson chambers, but statistics are extremely poor. There is, then, no evidence to suppose that the stars observed in the photographic plates are due to radiation which, while ionizing too little to affect the plate, would be visible in a Wilson chamber.

On the other hand, a large proportion of the collimated stars observed in the chamber are due to ionizing radiation. We have seen that these stars are at most a small percentage of the total. It seems that they are exceptional not only in their appearance, but also in the manner in which they are produced, and they may perhaps be due to fast protons or mesons. So far the photographic work has told us nothing about them, and the Wilson chamber experiments are not numerous enough to enable us to come to a definite conclusion.

If we exclude all kinds of ionizing radiation and photons on the grounds given above, we are left with three possible causes: neutrons, neutral mesons (of the kind suggested by Powell to explain the decay of his heavy mesons, and by Anderson to explain that of the normal cosmic ray meson), and finally some hitherto unknown particle. I exclude, of course, the few stars which are due to σ mesons.

There is little in the momentum-energy relations of the isotropic stars to give us a clue to their origin. The mean velocity of the particles emitted is of the order 5×10^9 cm./sec. We should hardly expect to detect an asymmetry caused by a motion of the centre of gravity of 3×10^8 cm./sec. (1 % of the velocity of light). The energy of the larger stars is of the order 500 MeV, and these stars are certainly due to nuclei at least as heavy as that of bromine. If a particle of rest-mass m, velocity V, is captured but not destroyed by a nucleus of mass M the velocity of the centre of gravity is $v = \frac{(E + mc^2)\,V}{(M+m)\,c^2}$, where E is the energy transferred from the particle. In our case $m \ll M$. If the particle is a *neutral meson* of mass $120m_e$, $V \sim c$ in order to give sufficient energy, $E \gg mc^2$ and $v/c \doteqdot E/Mc^2$; since E/Mc^2 does not exceed $\frac{1}{160}$, this gives an asymmetry which would be undetectable. If the rest mass of the meson is turned into energy there will be less demands on the kinetic energy and hence less momentum transferred, so the asymmetry will be even less detectable. If the meson goes through the nucleus without capture, the momentum for a given energy transfer is less than when capture occurs.

If we suppose that the particle is a *neutron*, $mc^2 \gg E$ and we can write approximately $v = Vm/M$. For a large star, which must necessarily come from a heavy nucleus, $m/M \ngtr \frac{1}{80}$, and V to give an energy ~ 100 MeV is $\sim 1 \cdot 5 \times 10^{10}$ cm./sec., so $v/c \ngtr \frac{1}{160}$, which would escape detection. Some of the small stars are no doubt due to light nuclei, e.g. carbon. Here $m/M = \frac{1}{12}$. If the energy is as much as 100 MeV, v/c would be $\frac{1}{24}$, which might be detectable by a careful statistical analysis.

For the collimated stars of the Wilson chamber, particle energies are higher. Many of the particles can penetrate a 1 cm. plate of lead which takes about 80 MeV and gives a velocity of the order of half that of light. To give the observed asymmetry the velocity of the centre of gravity must be of the same order. This velocity, when the incident particle is retarded to lose an energy ΔE, is approximately $\Delta E/MV$; V will be of the order of c and so $\Delta E/Mc^2 \sim \frac{1}{2}$. If the incident particle has the necessary energy there is nothing impossible in this. In these cases it makes little difference whether the incident particle sticks to the rump of the nucleus, if any, or goes on, since all velocities are of the order of that of light.

Against the hypothesis of neutral mesons is the fact that the rate of increase with altitude of the stars is much greater than that of the annihilation of the meson component of the cosmic rays, which would be inadequate to provide the number of stars observed at high altitudes. The number of heavy mesons undergoing change always appears to be small.

Most of the neutral mesons are formed by the decay of heavy and light charged mesons at rest, in which cases the neutral meson only gets a few million volts kinetic energy. If neutral mesons were the main cause of stars the energy of the stars would be constant, namely that due to the rest mass of the neutral meson, say 60 MeV. Unfortunately, we cannot see the neutrons which are no doubt associated with the stars, and this makes it hard to say if the observations are compatible with a constant total energy, but on the whole the character of the energy-frequency curve is against the hypothesis though not perhaps decisively so.

It is, of course, possible that the bulk of stars are due to neutral mesons coming from the top of the atmosphere with energy considerably exceeding their rest mass. If so, it is easy to calculate the cross-section for star production in an atom of nitrogen or oxygen. It comes to $1 \cdot 6 \times 10^{-25}$, a not impossible value. On this view one would assume that the neutral meson was not destroyed in the process, or at least not always.

Neutrons, on the other hand, show the same variation with altitude as the stars; they are present in considerable numbers. The methods which have been used up to the present to detect them would not be effective for particles with the high energy needed to produce stars, but it is not unreasonable to suppose that there may be some with great energy, that these energetic neutrons either produce stars and so disappear, or lose energy by collisions and appear among those which are detected by the other methods.

THE DISTRIBUTION OF THE STARS OVER THE SURFACE OF THE PLATES

The stars show the expected random distribution over large areas. Thus 1230 stars on 54 sq.in. of plate were distributed among areas of $1 \times \frac{1}{2}$ in. according to a Poisson distribution. When, however, the examination is pushed to smaller distances a curious effect appears. This has been discovered simultaneously by Leprince-Ringuet & Heydmann (1948) in France and by Li & Perkins here (1948). It consists in a tendency for pairs of stars to appear near each other more frequently than would be expected if they were distributed at random. The effect is marked for distances up to about 500μ, but for larger distances becomes hardly statistically significant. It is not at present possible to say whether this apparent limitation of the influence of one star upon another is real, or whether it simply arises from the mathematical fact that the number of stars on a thin plate within a distance r from a fixed point will increase as r^2 if the stars are at random, while those due to connexion with a star at the origin would probably vary as $\log r$ and could at most be expected only to increase as r. Thus the effect of any influence gets relatively smaller at large distances and is lost in the statistical fluctuations even though there may be no true absorption.

It ought to be possible to check whether the effect is due to a secondary radiation from the star or due to the production of twin stars by two of a sheaf of star-producing particles diverging from a common source, which would lead to a different law of variation with r of the excess of the number of stars over that expected from random distribution, but the present statistics are hardly good enough, though for what they are worth they favour the second hypothesis. Assuming for the sake of getting a measure of the effect that it is due to a radiation from the stars, or from some of them, producing secondary stars, we can calculate the product of the number of such particles per star and their average cross-section for star production in the atoms of the emulsion. It comes surprisingly high, 14×10^{-24} cm.^{2}. It seems almost impossible that there are several particles emitted per star with energy large enough to produce an apparently similar star, and the reasonable view would be that the emission of even one such particle is a tolerably rare event. But if this is the case the cross-section for star production must be several times the above value which is only explicable by some process of resonance.

If we suppose that the twin stars are produced by two members of a sheaf of hypothetical particles the cross-section required comes out a little less, but it is still surprisingly high. It is hoped that work with thicker films of emulsion, which should give more twins, will help us to clear up this very puzzling difficulty.

We are left then with a number of problems: What is the origin of the stars? Are they due to neutrons, to neutral mesons or some still-to-be-discovered particle? Are the alpha-particles emitted by a different process from the protons? Is fission a normal accompaniment of the process which forms a star? What is the explanation of the close pairs of stars? Is this a clue to the process of star formation in general or does it only concern a small proportion of the stars? All these questions should be answered in the next few years. Some of them we hope in the next few months.

REFERENCES

Agnew, H. M., Bright, W. C. & Froman, D. 1947 *Phys. Rev.* **72**, 203.
Bethe, H. A., Korff, S. A. & Placzek, G. 1940 *Phys. Rev.* **57**, 573.
Bagge, E. 1941 *Ann. Phys., Lpz.*, **39**, 512.
Blau, M. & Wambacher, H. 1937 *S.B. Akad. Wiss. Wien*, **146**, 469, 623.
Gardner, E. & Paterson, V. 1948 *Phys. Rev.* **73**, 533.
Halban, H., Kowarski, L. & Magat, M. 1939 *C.R. Acad. Sci., Paris*, **208**, 572.
Hazen, W. E. 1944 *Phys. Rev.* **65**, 67.
Heisenberg, W. 1938 *Nuovo Cimento*, **15**, 31.
Korff, S. A. 1939 *Phys. Rev.* **56**, 210.
Korff, S. A. & Hamermesh, B. 1946 *Phys. Rev.* **67**, 155.
Lattimore, S. 1948 *Nature*, **161**, 518.
Lattes, C. M. G., Occhialini, G. P. S. & Powell, C. F. 1947 *Nature*, **160**, 486.
Leprince-Ringuet, L., Heydmann, J., Li, T. T. & Perkins, D. H. 1948 *Nature*, **161**, 844.
Powell, C. F. & Occhialini, G. P. S. 1947*a* *Nature*, **159**, 93.
Powell, C. F. & Occhialini, G. P. S. 1947*b* *Nature*, **159**, 186.
Powell, W. M. 1946 *Phys. Rev.* **69**, 385.
Stetter, G. & Wambacher, H. 1938 *Phys. Z.* **39**, 883.
Stetter, G. & Wambacher, H. 1939 *Phys. Z.* **40**, 702.
Weisskopf, V. 1937 *Phys. Rev.* **52**, 295.
Yuan, L. C. L. 1948 *Phys. Rev.* **74**, 504.

Nuclear disintegrations produced by cosmic rays

By J. B. Harding, S. Lattimore and D. H. Perkins
Imperial College of Science and Technology, London

(*Communicated by Sir George Thomson, F.R.S.—Received* 24 *August* 1948—
Revised 5 *November* 1948)

Measurements have been made on the energy and angular distributions of the charged particles from disintegration 'stars' produced in the silver and bromine nuclei of photographic emulsions exposed to cosmic radiation. The observations extended over a wide range of excitation energies (100 to 700 MeV). The energy spectra and angular distributions of the protons can be explained in all cases by simple evaporation theory. This energy distribution shows also a high-energy tail consisting of direct knock-on protons and slow mesons. At high excitation energies the α-particles exhibit collimation effects which are probably due to localized 'boiling' or a form of fission.

1. Introduction

The first extensive observations on the multiple nuclear disintegrations or 'stars' produced by cosmic rays were made by Blau & Wambacher (1937). These and subsequent investigators, using the photographic emulsion technique, measured the increase of star frequency with altitude, and obtained distributions of the numbers and energies of the charged particles from stars. Cloud-chamber investigations of the cosmic ray disintegrations, especially those leading to the production of penetrating particles, were made by W. M. Powell (1946) and Hazen (1944). With the recent development of emulsions of greatly increased sensitivity and discriminating power, it has become possible to trace regularities in the modes of disintegration of nuclei under high excitation, and, in the present paper, we describe investigations on the cosmic ray stars produced in the Ilford 'Nuclear Research' emulsions.

In § 2 are discussed the recording characteristics of the emulsions used, and the methods of identifying the tracks of the various types of particles present in the disintegrations. Observations on the energy and angular distributions of the particles are described in § 3 and discussion of results in §§ 4 and 5.

Owing to the rapid increase of the star intensity with altitude, most of our observations were taken from plates exposed in mountain stations on the Jungfraujoch (3457 m.) and at Chamonix (3650 and 4400 m.).

2. General technique

(a) *Types of particles emitted from stars*

In examining a disintegration, it is of course desirable and sometimes essential to be able to identify the tracks of the various charged particles emitted. In processes involving particles of widely varying velocity, charge and mass, the recording characteristics of the emulsion used must essentially represent a compromise between high sensitivity on the one hand, and good discriminating power on the other. The reason for this is that when the specific ionization exceeds a certain critical value, the latent images in the silver bromide grains along the path of the ionizing particle are produced so close together that the silver streamers growing out of the grains during development run together, or 'clog', so as to form a more or less 'solid' track. It is then extremely difficult to perform a reliable 'grain-count' along the track, and discrimination between the various types of particles is poor. The higher the sensitivity of the emulsion, the lower is this critical value of the ionization. In our experiments, the plates used (Ilford B 2 and C 2) and development conditions were such that we could detect protons with energies up to about 100 MeV. Discrimination between even slow protons and fast α-particles (up to well over 50 MeV) was very good, but that between the solid tracks of α-particles and heavier nuclei was in general extremely difficult.

Usually the emitted particles have only a small fraction of their total range in the emulsion, and so the lightly ionized tracks of fast protons, mesons, deuterons, tritons and extremely fast α-particles cannot be distinguished unless at least half their true range is observed and a statistically significant number of grains counted.

In most of the stars, we observe (in addition to long tracks of mesons, protons, deuterons, tritons, α-particles and heavier nuclei) short, very heavy tracks produced by 'recoil fragments'. These tracks have nearly always a range of between 2 and 5 μ.

(b) *Grain-counting for energy determinations of lightly ionized tracks*

When the ionization is below the critical clogging value defined above, we can determine the velocity of a particle, if its charge is known, from the mean grain-density. This is obtained from a grain count and total track length. The latter quantity is determined from its horizontal projection as measured in the eyepiece scale of the microscope, and from depth measurements on the fine adjustment drum. The average thickness of the unprocessed emulsion on a given plate was assumed to be accurately the value given by the manufacturers (100 μ). The average thickness in drum divisions of the processed emulsion on a plate was measured, and equated to 100 μ. Thus conversion factors, varying slightly from plate to plate due to variations in the average thickness of processed emulsion caused by slight variations in washing and drying conditions, were obtained. These allowed the fine adjustment reading to be converted into true track depth in the unprocessed emulsion.

A grain-density/range curve has been constructed by using grain-counts along complete proton tracks, the longest having a range of about 4000 μ. This curve has been extended to 15,000 μ by extrapolating from a long meson track (over 2000 μ in length), taking into account the difference in mass between meson and proton.

The range-energy relations for α-particles and protons in Ilford unloaded emulsion have been determined experimentally by the Bristol group up to about 15 MeV. These curves have been extended up to velocities of 0·4c (100 MeV protons) using theoretical formulae for the collision loss (Rossi & Greisen 1941). In the computation we have assumed the Bloch formula $I = I_H Z$ for the mean ionization potential of an atom (I_H = ionization potential of hydrogen, Z = atomic no.); since this relation is only very rough, and there are uncertainties in the determination of the emulsion constitution, the error in the proton curve may be 10 % at 100 MeV.

From the range-energy and grain-density range curves the relation between grain-density and energy for protons is then easily obtained.

3. ENERGY AND ANGULAR DISTRIBUTION OF CHARGED PARTICLES

In detailed analyses of 78 nuclear disintegrations of silver and bromine, each disintegration comprising the emission of between 7 and 22 charged particles, and of 53 smaller disintegrations, the following procedure was adopted.

In order to study the variations in the modes of disintegration with excitation energy, the stars were divided into the four groups as shown in table 1.

TABLE 1

no. of tracks per star	no. of stars examined	average no. of protons per star	average no. of α-particles per star	ratio of alphas to protons	estimated mean excitation energy (MeV)
15 to 22	26	10·5	6·2	0·59 ± 0·06	650
10 to 14	29	5·7	4·4	0·77 ± 0·09	325
7 to 9	23	3·9	3·3	0·85 ± 0·13	200
3 to 6	53	2·3	1·5	0·65 ± 0·09	100

Column six was obtained by summing the kinetic and binding energies of the emitted alphas and protons, and adding an amount of energy equal to that associated with the protons to allow for the neutrons.

(*a*) *Proton, and other lightly ionized tracks*

The grain densities of all lightly ionized tracks were determined. In the majority of the stars a proton came to the end of its range in the emulsion, enabling one to construct a separate grain-density/energy curve for each disintegration. Assuming all lightly ionized tracks to be due to protons, unless grain-density and range measurements indicated tritons, etc., the energies, corrected for fading, were determined. The maximum differences of grain-density (for equal velocities) between most and least faded tracks were of the order of 10 % with the exposures used (40 days).

The energy distributions of the protons are indicated in the histograms of figure 1. The error in determining the proton energy from the grain-count should correspond to the statistical error arising from the finite number of grains counted.

The average percentage errors at various energies were as shown in table 2.

TABLE 2

E (MeV)	8	15	25	35	80
$\pm \Delta E$	1·3	2·45	4·45	7·6	20
percentage error (±)	16	16½	18	22	25

Owing to slight non-uniformity of mixing of the emulsions during manufacture, the actual errors are likely to be a little greater than those indicated.

In order to determine whether the slight variation of development with depth in the emulsion has any serious effect on the grain-density along the tracks, the apparent energy distribution of protons passing out of the emulsion-air surface has been compared with that of those passing into the glass (table 3, figures give number of tracks).

TABLE 3

energy (MeV)	surface	glass
0 to 5	9	12
5 to 10	61	68
10 to 20	57½	60
20 to 30	32½	26
30 to 50	18	21
> 50	11	4
total	189	191

No significant difference between the two distributions is observed (the plates were exposed vertically).

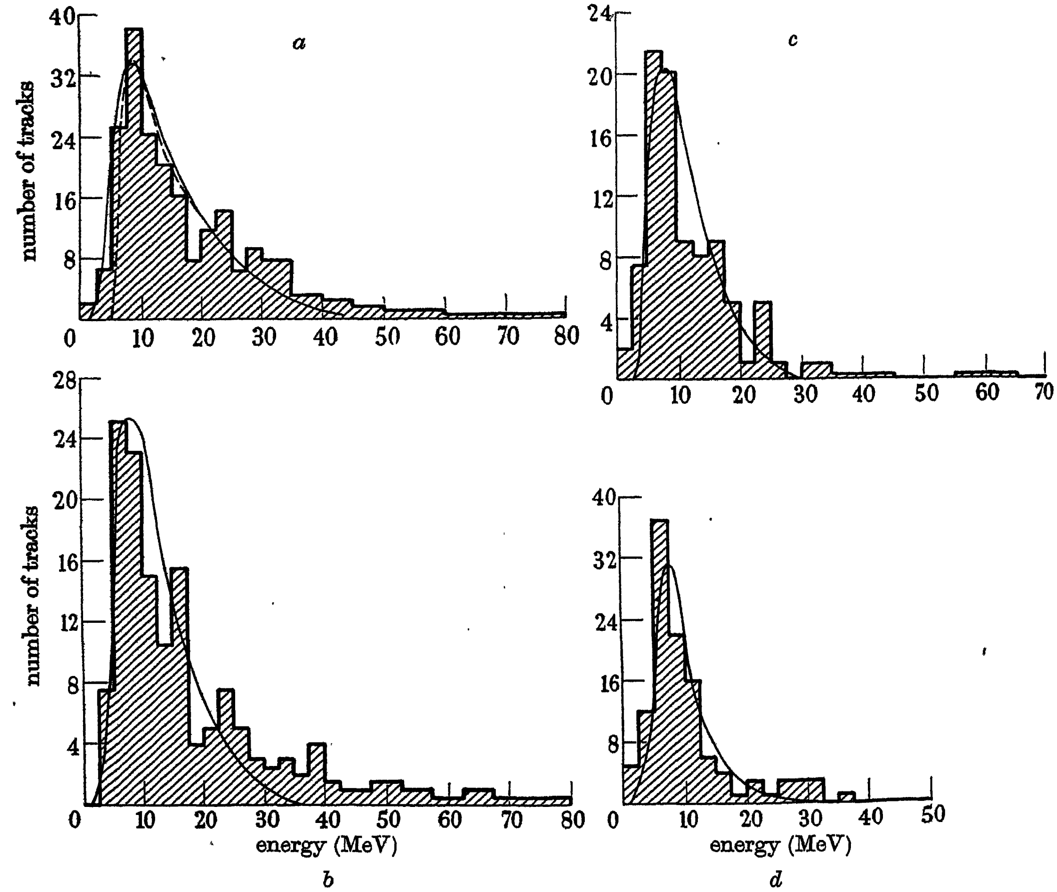


FIGURE 1. Energy distribution of proton tracks from stars with *a*, 15 to 22 tracks ($T_0 = 8$ MeV); *b*, 10 to 14 tracks ($T_0 = 5\frac{1}{2}$ MeV); *c*, 7 to 9 tracks ($T_0 = 4\frac{1}{2}$ MeV) and *d*, 3 to 6 tracks ($T_0 = 3$ MeV). Full curves correspond to evaporation formula, allowing for cooling and barrier penetration. Dashed curve corresponds to classical calculation (see § 4).

(*b*) *Energy distribution of α-particles*

Owing to the fact that the tracks of α-particles are practically 'solid' their energies cannot be determined by a grain-count (not at any rate if $E < 70\,\text{MeV}$). Hence the energy distribution can only be obtained by measurement of the ranges of those alphas ending in the emulsion.

The probability that an α-particle track, of length l and random direction, lies completely in the emulsion, is

$$f = \frac{d}{2l} \quad \text{for } l > d,$$

$$f = \left(1 - \frac{l}{2d}\right) \quad \text{for } l < d,$$

where d is the emulsion thickness. The energy distribution of complete α-tracks from stars with 10 to 22 charged particles is given in figure 2. This distribution is corrected for 'loss' by multiplying each ordinate by a factor $1/f$, and the result is indicated in figure 3. (Owing to the large value of f for the very long-range α-particles, the statistical error at the high-energy end is fairly large.)

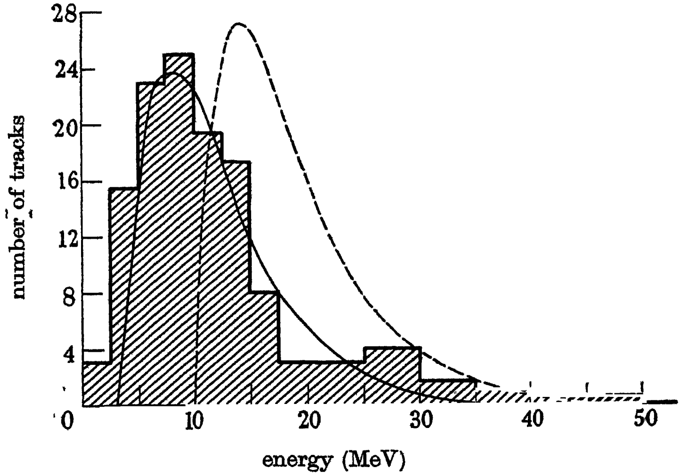


FIGURE 2. Energy distribution of complete α-tracks (stars with 10 to 22 tracks). Curves indicate the evaporation distributions for an instantaneous temperature of 7 MeV and barrier heights of 3 MeV (——) and 10 MeV (– – – –) (see § 5*a*). These have been multiplied by the correction factor f.

(*c*) *Angular measurements on protons and α-particles*

The complete orientation of a proton or α-particle in the emulsion was obtained from the azimuthal angle θ (measured in the eyepiece goniometer), and the angle of dip or latitude ϕ obtained by a depth measurement. The value of θ (allowing for a small amount of Coulomb scattering along the tracks) can be found to $\frac{1}{2}°$, but the angle of dip is not known so accurately, especially for large values of ϕ. Owing to the curvature of the field of the microscope objective, the difference in depths of a track at two points along it must be obtained from the depth values relative to a specific reference level (the surface of the emulsion). This surface is only approximately plane. Measurement of ϕ also depends on the linearity of shrinkage of the emulsion. Tests indicate that to within the limits of error (5 %) in setting and reading the fine-adjustment screw, the emulsion shrinkage during processing is linear.

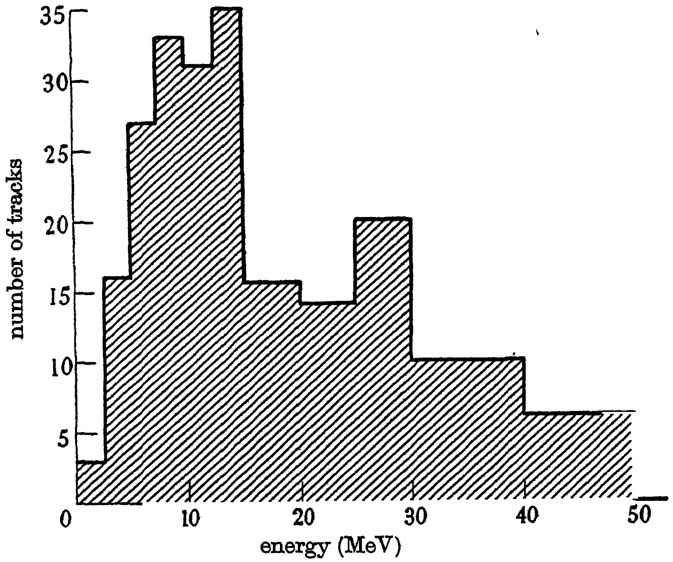


FIGURE 3. Energy distribution of complete α-tracks multiplied by correction factor $1/f$.

4. DISCUSSION OF RESULTS FOR PROTONS

(a) *Theoretical treatment of energy spectrum*

The energy distribution of heavy particles produced in stars in photographic emulsion was first investigated by Wambacher (1938) and later by Ortner (1940). In these measurements no particular regard seems to have been paid to the effect of latent-image 'fading' and the discrimination between protons and α-particles was very poor with the plates then available (Ilford halftone). Thus an attempt by Bagge (1941) to explain their results in terms of the 'evaporation model' of the nucleus did not appear particularly convincing.

The evaporation model of the nucleus was put forward by a number of authors, notably Bohr, Frenkel, Landau, Bethe and Weisskopf. Owing to the very close spacing of energy levels in a heavy, highly-excited nucleus, one can average over a large number of quantum states of nearly equal energy, and thus apply thermodynamical reasoning in order to determine the distribution of energy among the nuclear particles. The disintegration process is then analogous to the evaporation of a drop of liquid. The energy distribution of the nuclear particles emitted has been derived rigorously by Weisskopf (1937). When, as in the cases we are considering, the nucleus is highly excited, the probability of a neutron being emitted with energy between E and $E+dE$ is approximately of the Maxwellian form

$$P(E)\,dE = \frac{E}{T^2}\,e^{-E/T}\,dE. \tag{1}$$

Both the temperature T and energy E are measured in MeV. The temperature T is related to the excitation energy U according to $U = \beta T^n$, where β is a (dimensional) constant and n depends on the assumed interaction between the nuclear particles (gas model, liquid drop model, etc.).

In the case of charged particles, equation (1) can be shown to become

$$P(E)\,dE = \frac{E-V}{T^2}\,\mathrm{e}^{-(E-V)/T}\,dE \quad \text{for } E > V, \tag{2}$$

where V is the height of the potential barrier, assumed impenetrable. Thus the effect of the barrier is simply to shift the whole energy distribution by an amount V. In a previous note (Perkins 1947) it was shown that the energy spectrum of protons from the disintegration of heavy nuclei (silver and bromine) in the emulsion could be explained in a general way by (2). This formula would be true for the energy distribution of molecules evaporating from a drop of water, but it cannot be strictly applied to the multiple disintegration of a nucleus, for the number of particles (~ 100) is so small that the nucleus cools appreciably after emission of each particle. The effect of cooling was calculated as follows.

Suppose a nucleus of mass number A_0 is excited to an energy U and initial temperature T_0. The problem is to find the temperature T of a subsequent nucleus A. The average energy in distribution (2) is $2T + V$, that in (1) is $2T$. Hence the average energy of neutrons and protons evaporating from a nucleus at temperature T is $2T + \bar{V}$, where $\bar{V}$ is the effective potential barrier for a neutron-proton pair, and is $\sim \frac{1}{2}V$. The average energy taken away by a proton or neutron is

$$\frac{dU}{dA} = E_0 + 2T + \bar{V} = 2T + \alpha, \tag{3}$$

where $E_0 \sim 8$ MeV is the binding energy per nucleon. Considering the distribution of figure 1a (15 to 22 charged particles per star) the average kinetic energy of a proton is 15 MeV. The energy required to release a proton is thus $E_0 + 15 = 23$ MeV. The average number of protons per star is 11, giving 250 MeV for the total energy for ejecting the protons. Assuming the neutrons carry away approximately the same energy, the excitation energy required $U \sim 500$ MeV. It will be shown later (see § 5) that there are strong reasons for believing that most of the α-particles are ejected in a preliminary process, followed afterwards by evaporation of neutrons and protons. Allowing for this process, the average value of A_0 is estimated at 80. We have calculated the cooling effect according to the following well-known nuclear models.

Fermi-gas model

If we consider the nucleus as an assemblage of essentially free particles (i.e. kinetic energy $\gg$ interaction energy), then it is equivalent to a gas obeying Fermi statistics. The excitation energy of such a system is given by

$$U = KAT^2, \tag{4}$$

where A is the number of particles. The constant in (4) is $K = \pi^2/4\epsilon$, where ϵ is an average Fermi energy for neutrons and protons; according to Bethe (1937), ϵ should be about 12 MeV. Calculations by Bardeen indicate that the effect of nuclear interaction would be to about double this value. Thus we take $K = 0{\cdot}1$. Hence, from (4),

$$T_0 = \sqrt{\frac{U}{KA_0}} = \sqrt{\frac{500}{0{\cdot}1 \times 80}} \sim 8\,\text{MeV}.$$

The figure for K must be very rough, and one must regard it as fortuitous that this value fits the experimental data quite well (see figure 1). Again, from (4),

$$\begin{aligned} \frac{dU}{dA} &= 2KAT\frac{dT}{dA} + KT^2 \\ &= 2T + \alpha \quad \text{(from (3))}. \end{aligned}$$

Hence

$$\frac{dA}{A} = \frac{2KT}{\alpha + 2T - KT^2}\,dT. \tag{5}$$

The solution of this differential equation gives a relation between temperature T and mass number A of any successive nucleus left after emission of $(A_0 - A)$ particles:

$$\frac{A}{A_0} \approx \left\{\frac{KT_0^2 - 2T_0 - \alpha}{KT^2 - 2T - \alpha}\right\} \left\{\left(\frac{T_0 - x}{T - x}\right)\left(\frac{T - y}{T_0 - y}\right)\right\}^{\frac{1}{\sqrt{(1+\alpha K)}}}, \tag{6}$$

where

$$x = \frac{1}{K} + \frac{\sqrt{(1+\alpha K)}}{K}, \quad y = \frac{1}{K} - \frac{\sqrt{(1+\alpha K)}}{K}.$$

The curve connecting A and T, obtained from (6), is shown in figure 4a. At $T = 0$, one finds $A = 55$, this being the mass of the residual nucleus. Thus the number of particles emitted is $A_0 - A = 80 - 55 = 25$. The average number of protons observed was 11, leaving 14 neutrons. This figure appears reasonable. The mean potential barrier during the evaporation process (corresponding to $A = 65$ to 70) is 5 MeV for protons; thus the value of $\alpha = E_0 + \bar{V}$ taken in this calculation was 10 MeV (i.e. $\bar{V} = 2\frac{1}{2}$ MeV).

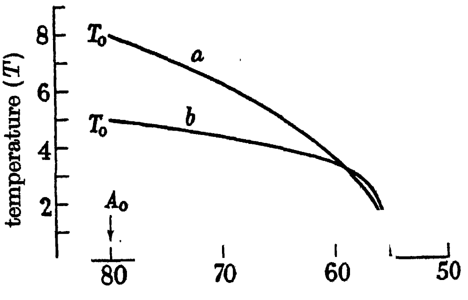


FIGURE 4. Relation between mass number A and temperature T of an evaporating nucleus according to a, Fermi-gas model $U = KAT^2$, $T_0 = 8$ MeV, and b, liquid drop model restricted to volume waves $U = KAT^4$, $T_0 = 5$ MeV.

From the T-A curve, values of T were obtained for 12 values of A (in steps of 2) and the energy distribution found at each temperature. The sum of these distributions calculated for $T_0 = 8$ MeV, and an average potential barrier V for protons of 5 MeV (corresponding to the average value of A) is shown in figure 1*a* (dashed curve). This spectrum shows quite good agreement with the experimental one, except at very low and very high energies.

The discrepancies at low values of energy can be explained by the quantum-mechanical penetration of the barrier. The penetration probability may be written $Q = Pl_c^2/l_0^2$, where $\hbar l_0$ is the maximum orbital momentum for which the particle interacts with the nuclear field ($l_0 = r/\lambda\!\!\!^-$, r being the nuclear radius, $\lambda\!\!\!^-$ the de Broglie wave-length of the particle). $\hbar l_c$ is an 'effective orbital momentum' obtained by summing up the penetration probabilities over all effective values of l, and P is the Gamow factor. Values of l_c^2/l_0^2 were calculated from the formulae of Bethe & Konopinski (1938). As usual $P = e^{-2g\gamma(x)}$, where $x = E/V$ (ratio of particle energy to barrier height) and the function $\gamma(x)$ has been tabulated by Bethe. For protons, $g = 0{\cdot}42Z^{\frac{2}{3}}$. In our case the mean values are $V = 5\,\text{MeV}$ and $g = 4$. We have attempted to allow for the penetration of the barrier by multiplying the distribution for uncharged particles (equation (1)) by the penetration factor $Q(E)$ for protons; the resultant distribution is given by the full-line curve (figure 1*a*). It is seen that for $E > V$ the agreement between quantum-mechanical treatment and the classical one is quite good. For $E < V$ the quantum-mechanical calculation shows reasonably good agreement with experiment.

Liquid-drop model

This model assumes the interaction between nucleons to be similar to that between molecules in a liquid (interaction energy of the same order as, or larger than, the kinetic energy). The excitation energy U may be expressed as (Bethe 1937)

$$\frac{U}{U'} = \tau^{\frac{7}{3}} + \tau^4.$$

The first right-hand term is due to surface, the second to volume waves. Here $\tau = T/T'$, where T' is the temperature at which contributions of energy U' from volume and surface waves are equal. The calculated value of T' for a nucleus of bromine or silver equals approximately 2 MeV. The calculated value for U' is 20 MeV.

For the excitation energies we are considering ($U \sim 500\,\text{MeV}$), the temperature $T \sim 4\,\text{MeV}$, and is therefore well above the critical temperature. Hence volume waves are more important than surface waves, so that U is approximately proportional to T^4.

The A/T cooling-curve for the liquid-drop (figure 4*b*) has a small gradient for $T \sim T_0$, and drops very steeply for small T. The difference between this curve and that for the Fermi-gas is due to the higher value of n in the formula $U = \beta T^n$. Thus even though the initial temperature T_0 in the case of the liquid drop (5 MeV) is much smaller than that for the Fermi-gas (8 MeV) it is found by calculation that the energy distribution from the liquid-drop model almost exactly coincides with that shown for the gas model in the region 0 to 15 MeV. Above 15 MeV, however, the liquid-drop curve drops off rapidly and shows very large discrepancy with the experimental distribution. (The reason for this is of course that the high-energy end of the distribution is determined largely by the initial temperature T_0.) Thus we conclude that, in so far as any 'model' of the nucleus may be described as correct, the Fermi-gas model appears to give the better results (or the liquid-drop model restricted to surface waves only, giving a dependence of energy on temperature according to a seven-thirds power law, not very different from that of the Fermi-gas model).

(*b*) *Comparison of experiment with theory*

Having fixed the constant K in equation (4) from the values of excitation energy and initial temperature T_0 for the stars with 15 to 22 charged particles, we proceed to calculate the initial temperatures for the other three groups of stars (table 4). Allowing for cooling as before, we obtain the distributions indicated by the full-line curves in figures 1*b*, *c*, *d*. It will be observed that the calculated energy spectra agree very well with the experimental histograms. It should, however, be pointed out that only about three-quarters of the stars involving 3 to 6 charged particles will correspond to disintegrations in silver and bromine nuclei. The remainder must be attributed to carbon, oxygen and nitrogen nuclei.*

TABLE 4

no. of charged particles N	estimated energy U required to eject protons and neutrons	T_0 (MeV)
15 to 22	500	8
10 to 14	250	5½
7 to 9	150	4½
3 to 6	80	3

In making this comparison of experiment with theory we must bear in mind the fact that a certain proportion of the so-called proton tracks have in fact been produced by deuterons and tritons passing out of the surfaces of the emulsion. From the stars examined we have observed 79 singly-charged particles coming to rest in the emulsion layer. 55 of these were identified as protons, and the remainder (apart from two slow mesons) as deuterons or tritons. Assuming the latter to be in equilibrium with protons and neutrons inside the nucleus, we expect the deuterons and tritons to have the same temperature and energy distribution as the protons. This implies that between 10 and 15 % of the so-called proton tracks which pass out of the emulsion surfaces are due to deuterons and tritons, for the latter together have about twice the chance of coming to rest in the emulsion as protons of similar energy. From this percentage of deuterons and tritons it is estimated that the energy values of the measured proton spectrum are about 7 % too low. The effect is thus quite small.

Thus we conclude that the energy distribution, between 0 and 30 MeV, of protons (and presumably neutrons) ejected from a heavy nucleus under very high excitation follows the form predicted by conventional evaporation theory, even though the temperature is continually dropping throughout the process, so that a state of thermodynamical equilibrium never really exists.

---

* The relative numbers of stars produced in the light and heavy nuclei have been determined by the use of plates consisting of alternate layers of normal emulsion and pure gelatin.

(*c*) *Angular distribution of protons*

In the course of an evaporation process the particles which are emitted should be orientated quite at random, i.e. the angular distribution should be spherically symmetrical with respect to the centre of mass system.

First of all it is necessary to estimate any effect on the angular distribution as a result of the momentum taken up by the whole (silver or bromine) nucleus from the incident particle. This momentum will be greatest, for a given energy transfer, if the star-producing particle actually stops in the nucleus. In the case of a neutron, for instance, transferring 500 MeV energy to the nucleus, this momentum would be about 700 MeV/*c*. Thus the velocity component of the whole nucleus would be only $\sim 0{\cdot}01c$, which is quite small compared with the average velocity of ejection of the star particles ($\sim 0{\cdot}2c$). The effect on the angular distribution is therefore negligible.

Thus if we ascribe a vector of unit length to each proton track, the resultant of all the vectors of the star should have a scalar magnitude R, the probability distribution $f(R)$ of which is determined from random-walk theory. If the number of vectors N is fairly large (>4), then the distribution is practically Gaussian (Chandrasekhar 1943):

$$f(R)\,dR \propto R^2\,\mathrm{e}^{-3R^2/2N}\,dR.$$

Since the number N of protons varies from star to star, the distribution becomes

$$f(R)\,dR \propto R^2 \sum_i m_i\,\mathrm{e}^{-3R^2/2N_i}\,dR,*$$

where m_i is the number of stars with N_i protons.

In figures 5*a* and *b* we have compared the theoretical and experimental distributions for the stars with 15 to 22 and 7 to 14 tracks. No significant difference is observed between the two, supporting the view that the protons evaporate.

(*d*) *High energy tail*

As can be seen from the histograms of figure 1, at energies greater than about 30 MeV the observed number of protons greatly exceeds that predicted by evaporation theory, and this high energy 'tail' is certainly even greater than is indicated, for the following reasons. Evidence has been given that the angular distribution of protons is isotropic. Hence if the frequency of proton tracks be plotted against the angle of dip ϕ we would expect cosine distributions. The histograms, figures 6*a* and *b*, indicate that practically no protons of energies less than 30 MeV escape observation, but a considerable proportion of steeply-dipping tracks of higher energy are not detected. Further, even under optimum conditions (tracks almost in the plane of the emulsion), we are unable to detect protons of energy greater than about 100 MeV. It has been suggested by Heisenberg that the first stage in the production of a multiple disintegration consists of the transfer of a large amount of energy from the incident cosmic-ray particle to a *few* nucleons situated over a small volume (possibly on the 'surface') of the nucleus. Some of these nucleons will immediately escape, while the remainder will produce a general 'warming up', by inter-nuclear collisions, of the entire nucleus, which subsequently cools down by evaporation.

---

* For values of $N_i < 4$, the exact form of the random-walk distribution was used.

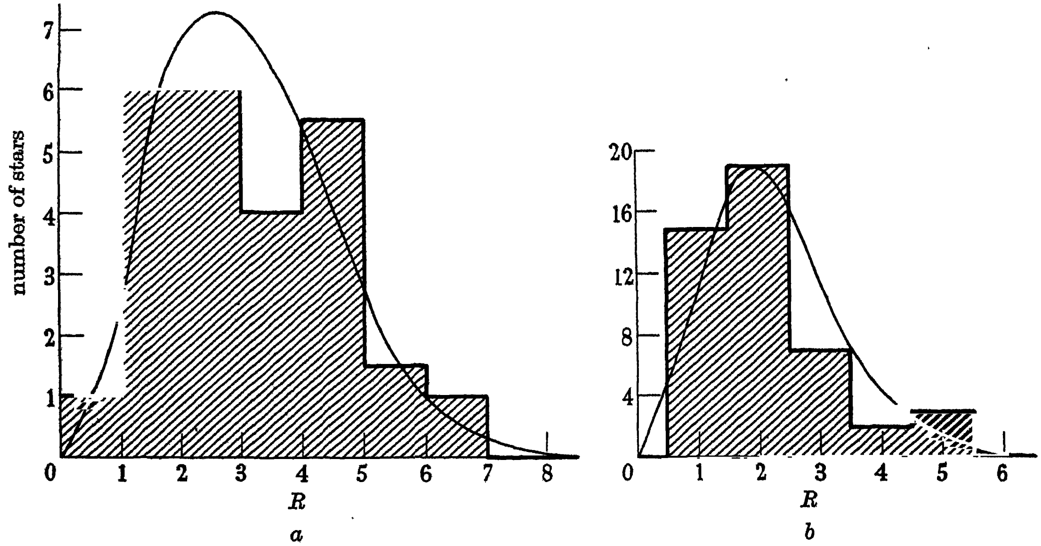


FIGURE 5. Frequency distribution of angular resultant R of protons from stars with *a*, 15 to 22 tracks; *b*, 7 to 14 tracks.

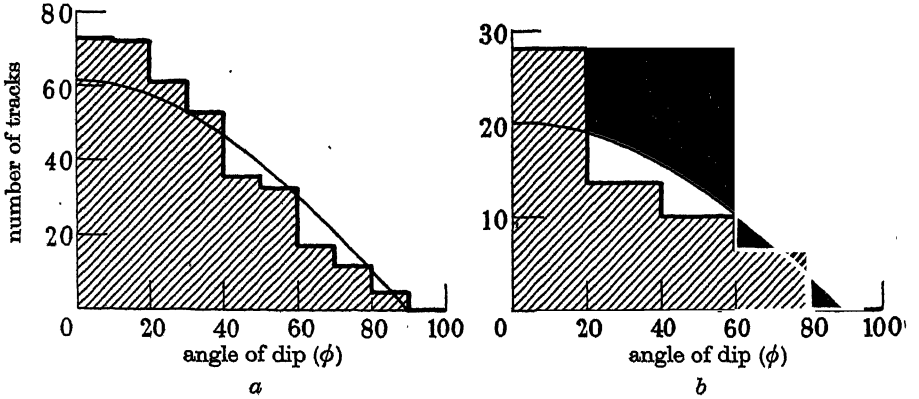


FIGURE 6. Dip distribution for proton tracks with energies *a*, below 30 MeV; *b*, above 30 MeV.

Clearly, the more nucleons knocked-on in the primary process by the incident cosmic-ray particle, the greater the ultimate excitation of the nucleus and the larger the star. Furthermore, the proportion of these nucleons which escape should be determined by the geometry of the nucleus alone, and should thus be more or less constant. Hence the number of 'Heisenberg' protons per star should go up roughly as the excitation energy, whereas, as shown in table 5, it increases more rapidly.

It seems highly probable, however, that a considerable proportion of the so-called fast protons in the Heisenberg tail are to be attributed to slow mesons. It is then natural to expect the rate of production of mesons to increase with the energy dissipated in the disintegration. Several examples of meson showers accompanied by heavy particles, i.e. nuclear fragments, have been observed in cloud chambers. The ejection of slow mesons from stars in photographic emulsions has been comprehensively treated by Powell & Occhialini (1948).

TABLE 5

'fast' protons per star (energy > 30 MeV)	excitation energy (MeV)
0·075 ± 0·025	100
0·175 ± 0·06	200
0·83 ± 0·15	325
1·42 ± 0·14	650

So far, from 55 stars with 10 to 22 tracks, we have observed two ejected mesons, of energies ~ 2 MeV, stopping in the emulsion. Thus although our statistics are extremely poor we have estimated, assuming isotropic ejection of mesons and a constant energy spectrum of production up to 10 MeV (maximum energy for a meson leaving a visible track in the emulsion), that between 10 and 30 incomplete meson tracks have been wrongly assigned as fast protons. This represents 15 to 45 % of the total number of particles in the tail of the proton distribution for these large stars.

It would be expected that fast knock-on protons should be collimated downwards, whereas slow mesons generated in nuclear disintegrations should be ejected isotropically. Preliminary investigations indicate downward collimation of lightly ionized tracks in the very small stars (3 to 6 tracks)—from 2000 of which we have so far observed only 1 ejected meson ending in the emulsion—but an isotropic distribution of these tracks in stars with more than 10 tracks (figure 7).

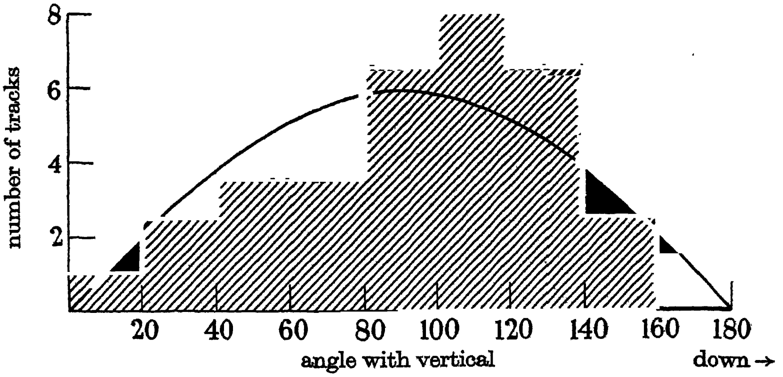


FIGURE 7. Zenith-angle distribution of lightly ionized tracks from stars with over ten tracks.

It will be clear from the above discussion that no very precise information is available regarding the nature of the particles producing the very lightly ionized tracks; it seems fairly safe to say, however, that at least half of them actually are fast protons.

(e) *Range of nuclear forces*

We interpret the protons in the high-energy 'tail' as those knocked out of a nucleus by direct impact of the energetic incident particle. Assuming this to be a nucleon, the energy distribution of the protons will be determined by the cross-section for neutron-proton (or proton-proton) scattering. Heisenberg (1938), Bagge (1944) and Rosenfeld (1948) have pointed out that this should lead to a rather precise determination of the range of nuclear forces. Assuming a Gaussian potential, Rosenfeld finds that the number $n(E)$ of protons with energies greater than E is approximately given by

$$\log_{10} n(E) = \text{const} - \left[0{\cdot}029\left(\frac{1}{xd}\right)^2 + 0{\cdot}0096\left(\frac{2r_0}{d}\right)^2\right] E,$$

where r_0 = nucleon 'radius', d = electron radius ($d \sim 2r_0$), and x^{-1} = range of force.

This equation holds only for the direct 'knock-on' protons, and hence will only apply to protons with energies in excess of about 30 MeV, where the contribution from 'evaporation' protons is negligible.

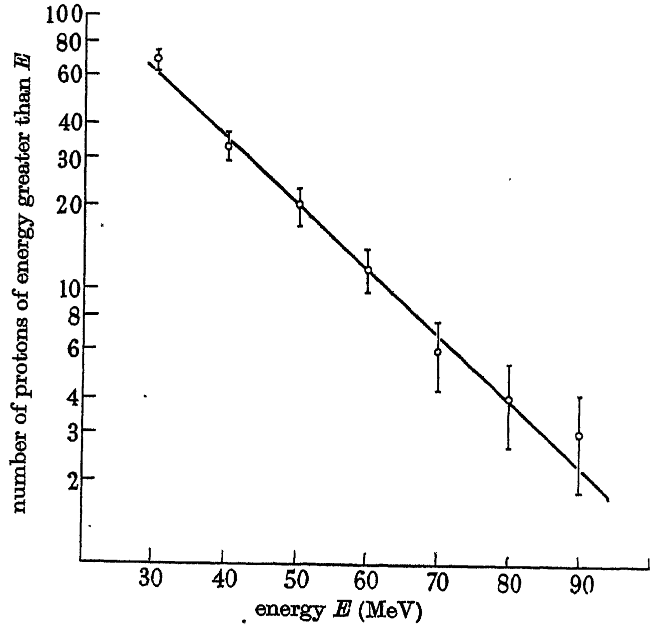


FIGURE 8. Determination of range of nuclear forces from the energy distribution of high-energy protons from stars.

The observed distribution, obtained from the stars with over seven tracks, is indicated in figure 8. A straight line drawn through the points has the equation $\log_{10} n(E) = \text{const} - (0{\cdot}0244 \pm 0{\cdot}004)\, E$. With $d = 2{\cdot}8 \times 10^{-13}$ cm. we obtain

$$x^{-1} = (2{\cdot}0 \pm 0{\cdot}3) \times 10^{-13}\,\text{cm}.$$

This value is in excellent agreement with those obtained by Bagge (1944) and Ortner (1940) by a similar method. It is also reasonably consistent with the range obtained by other methods.

We are, however, inclined to believe that this agreement is largely fortuitous, in view of the fact that the proportion of tracks escaping observation owing to their steep angles of dip increases with energy. On the other hand, the number of lightly ionized tracks may be overestimated at the higher velocities owing to presence of slow mesons in the so-called proton spectrum. Although these two effects will tend to cancel, we cannot attach much importance to the above figure for the range of the force.

5. DISCUSSION OF RESULTS OBTAINED FOR α-PARTICLES

(*a*) *Energy spectrum*

It can be seen from the histograms in figures 2 and 3 that the energy spectrum is considerably broader than the corresponding one for protons. The curves drawn in these figures are those of the evaporation distributions for *instantaneous* temperatures of 7 MeV and barrier heights of 3 and 10 MeV. In each case we observe an excess of α-particles of high energy (the *absolute* excess is quite large when the numbers have been multiplied by the factor $1/f$). Owing to the small number of tracks observed it is not possible to fit a temperature very accurately, but if the nucleus 'cools' appreciably during the emission of α-particles, a mean temperature of 7 MeV implies an initial temperature greater than that of the protons (5 to 8 MeV).

The most interesting feature of the energy distribution is that nearly half of the α-particles have energies *below* the potential barrier (about 12 MeV for the heavy nuclei in the emulsion). This would hardly appear conceivable if the α-particles are ejected according to an evaporation process. However, calculations by Bagge (1944) indicate that, at the excitation energies we are considering, the barrier might be reduced by as much as 40 or 50 % due to an increase in the effective nuclear radius caused by large amplitude surface waves. Such a reduction would at any rate partly explain the proportion of low energy alphas. The perturbations in the Coulomb barrier would be even greater if the α-particles are ejected as the result of intense localized heating which almost certainly occurs as a preliminary to the general 'warming up' of the whole nucleus. The low-energy end of the proton spectrum will, however, still be determined by those protons evaporating at low temperature, when the nucleus has cooled considerably, and the reduction in the barrier is therefore negligible; hence the considerations above will leave the energy distributions of the protons almost unaffected.

An alternative is to assume that when the excitation energy is of the same order of magnitude as the total binding energy, the possibility arises of fairly heavy 'splinters' breaking away from the parent nucleus. These 'splinters' would be very unstable and disintegrate spontaneously; the α-particles emitted would thus have to cross a much lower Coulomb barrier.

This phenomenon appears characteristic of strong nuclear excitation, for we have also made preliminary observations on the energy distribution of particles from the small stars (2 and 3 tracks). Here the average excitation energy is about 80 MeV, i.e. small compared with the binding energy of a silver or bromine nucleus. In this case we find that the α-particles evaporate from the nucleus and come out above the orthodox potential barrier.

(*b*) *Angular distribution of α-particles*

In figures 9 *a* and *b* are given the frequency distributions of the angular resultants for the α-particles and heavier nuclei (e.g. Li^8). For the stars with 7 to 14 tracks we obtain an isotropic distribution. This may indicate that at these excitation energies (i.e. up to half the total nuclear binding energy) the majority of the α-particles are emitted in an evaporation process. For those stars consisting of 15 to 22 tracks, however (figure 9*a*), we observe a deviation from the random-walk curve. The χ^2 test seems to show that this deviation is statistically significant; the probability that the observed distribution corresponds to spherical symmetry is less than 5 %. This asymmetry indicates that the α-particles ejected from heavy nuclei excited to energies ($\sim$700 MeV) comparable with the binding energy, exhibit collimation or 'bunching'.

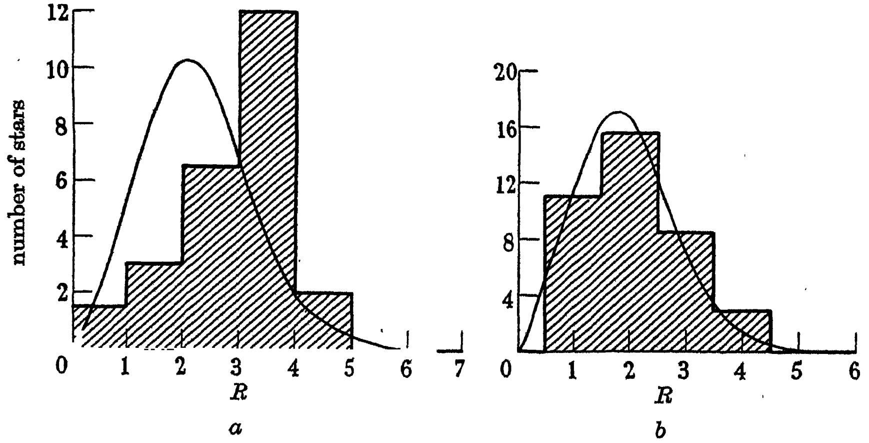


FIGURE 9. Frequency distribution of angular resultant *R* of α-particles from stars of *a*, 15 to 22 tracks; *b*, 7 to 14 tracks. Curves in all cases refer to random-walk distributions.

We have further investigated this collimation by measurements on the distribution in angle between the recoil fragment, or fragments, and the angular resultant of the protons or α-particles. We estimate that the time taken by the recoil nucleus to come to rest is at least 10^{-14} sec. On the other hand, the time required for the nucleus to reach 'thermal equilibrium' and subsequently evaporate will clearly be several orders of magnitude greater than the time between internuclear collisions (10^{-22} sec.). It seems fairly safe to say, however, that the evaporation process will be completed in less than 10^{-14} sec. In this case, the momentum taken up by the recoil fragment will be simply the vector sum of the momenta of the protons, neutrons and α-particles in the star, and the momentum contributed by the particle producing the star.

From the relative numbers of protons and α-particles, and their energy distributions, we are able to calculate the most probable resultant momentum P_{np} of the protons and neutrons (assuming equal numbers) and that of the α-particles P_α. Let Ψ be the angle between the fragment and angular resultant of α-particles or protons.

In the case of stars with 15 to 22 tracks, in which the alphas exhibit collimation (figure 9*a*), P_α is much greater than P_{np}, so that the direction of the fragment will be determined largely by that of the α-particle resultant. This is clearly indicated in figure 10*a*. On the other hand, the distribution for protons is more or less isotropic ($\sin\Psi$) (figure 10*b*).

Thus the two sets of measurements on the collimation effect are internally consistent. However, it must be emphasized that our data are obtained from only 26 disintegrations, and we should require to observe a much larger number of stars in order to establish the effect with complete certainty.

We can account for this asymmetry on either of the two hypotheses given to explain the number of low energy α-particles (§ 5*a*).

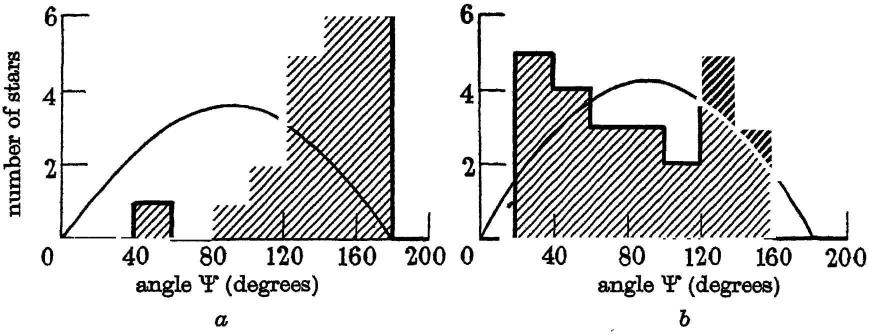


FIGURE 10. Distribution in angle Ψ between fragment and angular resultant of *a*, α-particles; and *b*, protons. Stars with 15 to 22 tracks. Curves indicate the isotropic distribution $\sin\Psi$.

First, we may imagine the α-particles to be ejected as a result of intense local heating near the surface of the nucleus. Those α-particles moving towards the surface will escape since their chance of colliding with other nuclear particles is small; on the other hand, those travelling away from the surface would have to traverse a large fraction of the nucleus and would therefore be stopped by collisions. One naturally expects the local heating to be more pronounced at the higher excitation energies, so that we are able to account for the transition from symmetry to asymmetry as the size of the star increases.

Secondly, we have the fission hypothesis in which a large fragment breaks away from the heavily excited nucleus. By virtue of the electrostatic repulsion such a 'splinter' would acquire considerable momentum. We must then postulate that the smaller fragment spontaneously disintegrates, preferentially into α-particles or heavier nuclei (e.g. Li^8), which thereby exhibit a strong collimation effect. The reason why α-particles, rather than protons and neutrons, should be emitted is difficult to explain, but hardly more so than that far more protons and neutrons are emitted as components of α-particles in our stars, than as nucleons themselves (table 1).

Evidence in favour of 'fission' comes from two sources. In the first place, out of the 26 stars we are considering, we obtain 4 certain cases in which two recoil fragments instead of one, are obtained. Secondly, we sometimes observe, especially in the very energetic disintegrations, examples of the concentration of a few nuclear particles in a very narrow cone, and of the emission of very heavy nuclear 'splinters' of long range.

According to these hypotheses, the rump nucleus should take up the momentum of the α-particles. The velocity component thus acquired by the protons and neutrons which subsequently evaporate from the rump will be negligible compared with their velocity of emission, so their angular distribution should be isotropic, as observed. When the rump nucleus has 'cooled down' completely, a recoiling fragment will be left behind. From the charge carried away by the protons and α-particles, we estimate the mass of the fragment at 25 to 45 units. The range of the recoil fragment is generally $\sim 3\mu$; we can thus get a rough idea of its momentum. It is of the order of magnitude of what we would expect from conservation of momentum of the α-particles; an exact calculation is not possible because of the unknown momentum contribution of the incident cosmic-ray particle (this may be of the same order of magnitude).

Since about 45 % of charged particles emitted from the stars are α-particles (table 1) a complete disintegration of bromine should give rise to about 25 tracks and of silver to 34. We have observed only two cases of such energetic stars in this laboratory; one of 23, the other of 25. Neither showed any sign of a recoil fragment. At present, therefore, we have insufficient large stars to obtain information on the process of disintegration of a heavy nucleus when the excitation energy is large compared with the nuclear binding energy.

Conclusions

Our conclusions regarding the production of a multiple disintegration in a silver or bromine nucleus may therefore be summarized as follows.

A high-energy cosmic-ray particle is incident on the nucleus, and, in its passage through it, collides with several nucleons. If the velocities of impact of these nucleons are directed suitably towards the nuclear surface they will escape immediately and produce a high-energy 'tail' in the energy distribution, as described by Heisenberg. If the momentum transfer between incident and scattered particle is sufficiently great, mesons may also be produced. Most of the recoiling nucleons pass further into the nucleus, and by collisions give rise to intense localized heating over a small volume of the nucleus. α-particles (and a few neutrons and protons) are ejected in the ensuing 'boiling off' process, which may also possibly give rise to fission, in which fairly large fragments break away from the main nucleus and disintegrate spontaneously into α-particles. The whole of the above process may be assumed to occur in a time of the order of magnitude of the nuclear transit time ($\sim 10^{-22}$ sec.). The excitation energy is then distributed throughout the remaining nucleus by numerous collisions, and this recoiling nucleus proceeds to evaporate off neutrons and protons (and a few α-particles) more or less isotropically. As the nucleus cools down, neutrons are predominantly ejected owing to the potential barrier. The evaporation process is carried out under quasi-equilibrium and is relatively slow (several orders of magnitude greater than the transit time). Any nuclear remnant left over gives a short recoil track. The last stage in the cooling consists of γ-ray and probably also β-transitions, neither of which can be observed.

We should like to express our sincere thanks to Professor L. Leprince-Ringuet for exposing our plates in his mountain observatory at Chamonix; and to Professor A. von Muralt and Dr R. Stämpfli for providing facilities at the Jungfraujoch Scientific Station. We are also indebted to Sir George Thomson for valuable discussions and advice during the course of the work.

References

Bagge, E. 1941 *Ann. Phys.* **39**, 512.
Bagge, E. 1944 *Phys. Z.* **21**, 461.
Bethe, H. A. 1937 *Rev. Mod. Phys.* **9**, 70.
Bethe, H. A. & Konopinski, E. J. 1938 *Phys. Rev.* **54**, 130.
Blau, M. & Wambacher, H. 1937 *S.B. Akad. Wiss. Wien*, **146**, 623.
Chandrasekhar, S. 1943 *Rev. Mod. Phys.* **15**, 1.
Hazen, W. E. 1944 *Phys. Rev.* **65**, 67.
Heisenberg, W. 1938 *Nuovo Cimento*, **15**, 31.
Ortner, G. 1940 *S.B. Akad. Wiss. Wien*, **149**, 259.
Perkins, D. H. 1947 *Nature*, **160**, 299.
Powell, C. F. & Occhialini, G. P. 1948 *Nature*, **162**, 168.
Powell, W. M. 1946 *Phys. Rev.* **69**, 385.
Rosenfeld, L. 1948 *Theory of nuclear forces*, Vol. I. Amsterdam: North Holland Publishing Company.
Rossi, B. & Greisen, K. 1941 *Rev. Mod. Phys.* **13**, 241.
Wambacher, H. 1938 *Phys. Z.* **39**, 883.
Weisskopf, V. 1937 *Phys. Rev.* **52**, 295.

A resonating-valence-bond theory of metals and intermetallic compounds

By L. Pauling, For.Mem.R.S.

The Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena

(*Received* 26 *July* 1948)

The resonating-valence-bond theory of metals discussed in this paper differs from the older theory in making use of all nine stable outer orbitals of the transition metals, for occupancy by unshared electrons and for use in bond formation; the number of valency electrons is consequently considered to be much larger for these metals than has been hitherto accepted. The metallic orbital, an extra orbital necessary for unsynchronized resonance of valence bonds, is considered to be the characteristic structural feature of a metal. It has been found possible to develop a system of metallic radii that permits a detailed discussion to be given of the observed interatomic distances of a metal in terms of its electronic structure. Some peculiar metallic structures can be understood by use of the postulate that the most simple fractional bond orders correspond to the most stable modes of resonance of bonds. The existence of Brillouin zones is compatible with the resonating-valence-bond theory, and the new metallic valencies for metals and alloys with filled-zone properties can be correlated with the electron numbers for important Brillouin polyhedra.

INTRODUCTION

The difficult problem of the electronic structure of metals has in recent years been tackled mainly by molecular-orbital (one-electron-orbital) methods. This attack has led to a reasonably satisfactory understanding of some metallic properties, such as electrical conductivity, but not of others, including interatomic distances (density), hardness and other mechanical properties, and the relative stability of alternative atomic arrangements. A decade ago (1938) I pointed out that the resonating-valence-bond concept of metals permits a straightforward explanation to be given not only of the hardness, melting-point, and related properties of the transition metals but also of their ferromagnetic saturation moments, which lead to the assignment of metallic valences such as 5·78 for iron and 5·44 for copper, in place of the numbers 0·2 to 0·6 for iron and 1 for copper assumed in the usual molecular-orbital treatment. It was suggested that these valence numbers result from the hybridization of the nine stable outer orbitals (five nd, one $(n+1)s$, and three $(n+1)p$) of a transition atom into 5·78 good bond orbitals (resembling the six d^2sp^3 orbitals of the central atom in an octahedral complex), 2·44 stable atomic orbitals (largely d in character, and responsible for the magnetic moment), and 0·78 'unstable' orbital, with no obvious use.

In the course of the further investigation of resonating valence bonds in metals the nature and significance of this previously puzzling 'unstable' orbital have been discovered, and it has become possible to formulate a rational theory of metallic valence and of the structure of metals and intermetallic compounds.

THE FUNDAMENTAL STRUCTURAL FEATURE OF A METALLIC SYSTEM: THE 'METALLIC ORBITAL'

Let us consider lithium as an example. In the usual treatment of this metal a set of molecular orbitals is formulated, each of which is a Bloch function built from the 2s orbitals of the atoms, or, in the more refined cell treatment, from 2s orbitals that are slightly perturbed to satisfy the boundary conditions for the cells. These molecular orbitals correspond to electron energies that constitute a Brillouin zone, and the normal state of the metal is that in which half of the orbitals, the more stable ones, are occupied by two electrons apiece, with opposed spins.

A roughly equivalent valence-bond theory would result from allowing the 2s electron of each lithium atom to be involved in the formation of a covalent bond with one of the neighbouring atoms. The wave function for the crystal would be

$$\psi = \sum_{i=1}^{M} a_i \psi_i, \tag{1}$$

in which†

$$\psi_i = 2^{-\frac{1}{2}N} \sum_{R} (-1)^R \left[\{(2N)!\}^{-\frac{1}{2}} \sum_{P} (-1)^P \, Pa(1)\,\beta(1)\,b(2)\,\alpha(2)\ldots \right] \tag{2}$$

is the wave function representing one of the M ways of distributing the valence bonds among the interatomic positions in the metal such that each atom forms one bond.

---

† See Slater (1931); for the meaning of the symbols see Pauling & Wilson (1935).

If the functions ψ_i are equivalent and orthogonal a_i will equal $M^{-\frac{1}{2}}$, and the energy of the normal state can be written as

$$W = W_{\text{bond}} + W_{\text{resonance}}, \tag{3}$$

in which $W_{\text{bond}} = \int \psi_i^* H \psi_i \, d\tau$ is the energy of any one of the valence-bond structures ψ_i and

$$W_{\text{resonance}} = \sum_{i=1}^{M} \sum_{j \neq i} a_i a_j K_{ij} = 2 \sum_{j=2}^{M} K_{1j} \tag{4}$$

is the resonance energy, K_{ij} being the resonance integral $\int \psi_i^* H \psi_j \, d\tau$ between two valence-bond structures i and j.

It is this resonance energy that would be in the main responsible for the difference in energy of the crystal and the gas of diatomic molecules Li_2. But the heat of formation of Li_2 molecules from atoms is only 6·6 kcal./g.-atom, whereas that of the metal is 39 kcal./g.-atom. It seems unlikely, by comparison for example with the analogous case of Kekulé-like resonance in aromatic molecules, that the great difference, 32·4 kcal./g.-atom, could result from the synchronized resonance, of type $\left\{\begin{array}{lll} \text{Li—Li} & \text{Li} & \text{Li} \\ & | & | \\ \text{Li—Li} & \text{Li} & \text{Li} \end{array}\right\}$, requiring two or more bonds to shift simultaneously. A much greater amount of resonance stabilization† can be achieved by resonance of the sort $\left\{\begin{array}{ll} \text{Li—Li} & \text{Li—Li}^- \\ & \quad\;\;| \\ \text{Li—Li} & \text{Li}^+ \; \text{Li} \end{array}\right\}$, in which one bond resonates independently from one position to another, through the process of electron transfer (Pauling 1948*a*).

However, this uninhibited resonance, involving the conversion of atoms into ions (or the transfer of ionic charges to atoms), *requires that the atom receiving a bond have an orbital available for its reception. It is the possession of this extra orbital* (*the metallic orbital*), *in addition to the orbitals required on the average for occupancy by unshared*

---

† A simple calculation may serve to illustrate this point. Let the $2N$ atoms in a lithium crystal containing N bonds (atomic arrangement A 2) be divided into classes A and B, at 000 and 1/2 1/2 1/2, respectively, in the unit cube. Let each atom be forming one bond. The total number of structures representing the eight bond positions for each atom A is 8^N. But only the fraction $(7/8)^7$ of these are, on the average, acceptable to a given atom B (the chance of its receiving a bond in any one position and not in others being $8/8\,(7/8)^7$); and hence the number of ways of placing bonds between adjacent atoms in the A 2 structure is

$$8^N (7/8)^{7N} = (7^7/8^6) = 3{\cdot}14^N.$$

The number 3·14 is a measure of the coefficient of the resonance integral for synchronous resonance.

If Li^+ and Li^- ions (the latter bicovalent) are also present, their *a priori* probabilities in class A are 1 and 28, respectively, with geometrical mean $2{\cdot}7^{\frac{1}{2}}$ (the ions must be present in pairs), which corresponds to 8 for neutral atoms. A calculation similar to that above, on the assumption that there is no energy difference between Li Li and $Li^+ Li^-$, leads to $(7^7/8^6)^N (1 + 7^{\frac{1}{2}}/2)^N$ for the number of ways of placing the bonds and hence to the number $(7^7/8^6)(1 + 7^{\frac{1}{2}}/2) = 3{\cdot}14 \times 2{\cdot}32$ as the measure of the coefficient of the resonance integral for uninhibited resonance. This result, containing the factor 2·32, indicates the importance of uninhibited resonance.

The relative weights $8 : 2{\cdot}7^{\frac{1}{2}}$ for Li and Li^+ ($= Li^-$) correspond to 43 % neutral atoms and 28·5 % each of anions and cations. Calculations based on minimizing the energy of the system also indicate similar values for the ratio of atoms to anions and cations. The rule of essential electrical neutrality (Pauling 1948*b*) would require that only the singly-charged ions need be considered.

electron pairs and bonding electrons, by all or many of the atoms in a condensed phase that permits the uninhibited resonance of valence bonds that gives rise to metallic properties.

HYBRIDIZATION OF BOND ORBITALS IN THE ALKALI METALS

In the molecule Li_2 the bond involves a hybrid atomic orbital $as+bp$ formed from the $2s$ orbital and one of the much less stable $2p$ orbitals. It is shown below that the amount of p character of this bond orbital (equal to b^2, with $a^2+b^2=1$) is small, being about 8 %. On the other hand, if each of the atoms in metallic lithium requires a bond orbital and a metallic orbital and the two are equivalent they will be $2^{-\frac{1}{2}}(s+p)$ and $2^{-\frac{1}{2}}(s-p)$, with 50 % p character. The analysis of energy quantities supports this conclusion.

It has been suggested (Pauling 1931; Pauling & Sherman 1937) that the bond-forming power of an orbital can be taken as proportional to its concentration in the bond direction, as measured by the value of its radial factor (normalized to 4π). The square of this quantity, the bond strength S, is approximately proportional to the bond energy. This postulate is equivalent to assuming that the off-diagonal matrix element H'_{sp} for $s-p$ hybridization is equal to $S_s S_p$, the diagonal elements being $H_s^0-\kappa S_s^2$ and $H_p^0-\kappa S_p^2$, in which H_s^0 and H_p^0 are the corresponding s and p atomic energies for the two bond electrons and κS_s^2 and κS_p^2 are the bond energies. S_s and S_p have the values 1 and $3^{\frac{1}{2}}$, respectively, and κ is a proportionality factor. The bond energy for formation of the most stable hybrid bond is then given by the lowest root (with changed sign) of the secular equation

$$\begin{vmatrix} -\kappa-W & 3^{\frac{1}{2}}\kappa \\ 3^{\frac{1}{2}}\kappa & H_p^0-3\kappa-W \end{vmatrix}=0, \tag{5}$$

in which H_s^0 has been taken as the origin for energy. Solution of this equation gives

$$-\frac{W}{\kappa}=\frac{1}{2}\left[\left\{\left(\frac{H_p^0}{\kappa}\right)^2-4\frac{H_p^0}{\kappa}+16\right\}^{\frac{1}{2}}+4-\frac{H_p^0}{\kappa}\right]. \tag{6}$$

The quantities $-W$ (the bond energy D_e of the Li_2 molecule) and H_p^0 (twice the difference in energy of the states $2p\,{}^2P$ and $2s\,{}^2S$ of Li) are known from spectroscopic data, and hence their ratio $\rho=-W/H_p^0$ is known. It is found from equation (6) that

$$\kappa=H'_s=\frac{\rho(\rho+1)}{4\rho+1}H_p^0. \tag{7}$$

The ratio b/a is the ratio of matrix elements in one row of the secular equation:

$$\frac{a}{b}=-\frac{\kappa+W}{3^{\frac{1}{2}}\kappa}. \tag{8}$$

The energy values and the derived quantities for the diatomic molecules of the alkali metals are given in table 1. It is seen that the amount of p character is calculated to lie between 5 and 14 %.

TABLE 1. BOND ENERGIES AND HYBRIDIZATION OF BOND ORBITALS IN DIATOMIC MOLECULES OF ALKALI METALS

molecule	$D_e = -W$	H_p^0	ρ	$\kappa = H_s'$	b/a	amount of p character (%)
Li_2	26·8†	85·0†	0·316	15·6†	0·41	14·0
Na_2	17·9	97·0	0·185	12·2	0·27	6·8
K_2	11·8	74·6	0·158	8·4	0·24	5·5
Rb_2	11·2	73·0	0·153	8·0	0·23	5·0
Cs_2	10·4	67·1	0·155	7·4	0·24	5·5

† The energy quantities in columns 2, 3 and 5 are in kcal./mole of diatomic molecules.

An independent estimate of the amount of p character of these bonds can be made with use of the assumption that a linear extrapolation of the low-lying vibrational energy levels (as indicated by the Morse potential function) will lead to the energy level of the atomic state involved in the bond. The equation

$$D_0^* = \frac{\omega_e^2}{4\omega_e x_e} - \tfrac{1}{2}\omega_e \tag{9}$$

for the linearly extrapolated energy of dissociation leads with use of the spectroscopic values for ω_e and $\omega_e x_e$ to the values of D_0^* given in column 4 of table 2. The ratio of the difference between this value and the actual energy of dissociation D_0 into atoms in the normal spectroscopic state and the p-s separation for two atoms (H_p^0, column 3 of table 1) is then the amount of p character in the bond orbitals. The values found in this way lie between 7·2 and 8·6 %, and are thus in approximate agreement with those found by the first method.

TABLE 2. HYBRIDIZATION OF BOND ORBITALS AS INDICATED BY LINEAR EXTRAPOLATION OF VIBRATIONAL ENERGY LEVELS

molecule	ω_e	$\omega_e x_e$	D_0^*	D_0	$D_0^* - D_0$	amount of p character (%)
Li_2	351·4	2·59	33·6†	26·3†	7·3†	8·6
Na_2	159·2	0·726	24·7	17·7	7·0	7·2
K_2	92·64	0·354	17·1	11·7	5·4	7·3
Rb_2	57·8	0·14	17·0	11·1	5·9	8·1
Cs_2	41·99	0·08	16·0	10·4	5·6	8·3

† In kcal./mole.

It is interesting to see, by comparison of the values in column 2 of table 1 (actual bond energy) and column 5 (calculated energy of pure s bonds) that this small amount of p character increases the bond energy by as much as one-half.

The results of the application of equations (7) and (8) to the metals themselves are given in table 3. The assumption is made that the bond-resonance energy depends on the bond strength S in the same way as the bond energy itself. The bond energy W is replaced by the heat of sublimation W_c. It is found that the calculated amount of p character is much greater than for the diatomic molecules. The calculated values, lying between 26 and 41 %, are, however, not so large as the 50 %

expected for the simple model discussed above. It seems likely that, whereas the two bond orbitals in the bicovalent anions M^- have 50 % p character, the occupied bond orbital in the neutral atoms has less than 50 % p character and the metallic orbital more. This difference in hybridization would decrease the value of the resonance integral but increase the atomic energy, and would result in the usual compromise that minimizes the total energy of the system.

TABLE 3. RESONANCE ENERGY AND HYBRIDIZATION OF BOND ORBITALS IN ALKALI METALS

metal	$-W_c$	ρ_c	κ_c	κ_c/κ	b/a	amount of p character (%)
Li	78·0†	0·918	32·0‡	2·05	0·83	41
Na	51·8	0·533	25·3	2·07	0·61	27
K	39·6	0·531	19·5	2·32	0·60	26
Rb	37·8	0·518	18·7	2·34	0·59	26
Cs	37·6	0·560	18·1	2·45	0·62	28

† Heat of sublimation in kcal./mole of bonds (2 g.-atoms of metal).

‡ Calculated energy of resonating s bonds in kcal./mole.

The ratio κ_c/κ given in column 5 of table 3 is the ratio of the energy of resonating s bonds to static s bonds. The values found, lying between 2·05 and 2·45, indicate that the energy of metallic resonance is indeed great, being as large as the bond energy (for fixed bonds). The increase by a factor of about 3 in the energy of the system accompanying the condensation of diatomic molecules to the crystalline metal can be divided into two factors; one of these, approximately 2, is due to resonance energy, and the other, approximately $\frac{3}{2}$, is due to change in hybridization, which is itself, to be sure, the result of the resonance energy in causing the p-s separation to bear a smaller ratio to the (resonating) bond energy.

The results of a simpler calculation for the alkaline-earth metals are given in table 4. The second column ($-W_c$) contains the heats of sublimation to atoms in the normal state $s^2\,{}^1S$. If the atoms are bicovalent, however, their hypothetical bond-forming state would be the configuration sp, with energy $\frac{1}{4}\{3({}^3P)+({}^1P)\}$ (that is, the weighted mean of the triplet and singlet states for the configuration sp), in case that the occupied bonding orbitals had 50 % p character (the metallic orbital being a pure p orbital), or would be a state with somewhat higher energy, in case that the metallic orbital had some s character, as in the alkali metals. If it is assumed, by analogy with the alkali metals, that the p character of the occupied bond orbitals lies midway between the values 50 and $66\frac{2}{3}$ %, corresponding respectively to a pure p metallic orbital and a metallic orbital equivalent to the occupied bond orbitals, the bond-forming state has energy $7/6H_p^0$, and the bond energy relative to this state becomes $7/6H_p^0-W_c$. Values of this quantity are given in the fourth column of the table.

It is interesting that the bond energy relative to the bond-forming state of the atoms shows the same monotonic trend for the alkaline-earth metals as for the alkali metals. The irregularity in the heats of sublimation at magnesium is due to the high energy of the bond-forming state of the atom, and not to any extraordinary property of the bonds themselves. The much larger bond energies of the alkaline-earth metals (by a factor of 2 for Be/Li and about 3 for the other pairs) is due mainly to a difference in character of the hybrid bond orbitals. If the alkaline-earth bond energies are divided by 3·88, the value of S^2 for *sp* bond orbitals with 41·7 % *s* character, as assumed above, there are obtained the values for the energy of resonating *s* bonds given in column 5 of table 4. The next column contains values of the ratio of this quantity κ_c' to the corresponding quantity for the alkali metals. These values are reasonable, in comparison with the ratio of single-bond radii— 1·38 for Be/Li, but somewhat smaller for the other pairs.

TABLE 4. ENERGIES OF RESONATING BONDS IN ALKALINE-EARTH METALS

	$-W_c$	H_p^0	$-W_c+\frac{7}{6}H_p^0$	κ_c'	κ_c'/κ_c
Be	75†	79·4‡	168‡	43‡	1·34
Mg	36·3	83·5	133·7	34·5	1·36
Ca	47·8	58·0	116·5	30·0	1·54
Sr	47	55·2	111	28·6	1·53
Ba	49	48·1	105	27·3	1·51

† The heat of sublimation in kcal./mole. ‡ In kcal./mole.

METALLIC RESONANCE AND METALLIC VALENCY

The straightforward way in which metallic valency can now be discussed may be illustrated by the example tin, which is more versatile in its behaviour than its congeners germanium and lead).

Tin has fourteen electrons outside of its krypton-like core. These may occupy the nine orbitals in the following three most stable ways (atomic electrons are indicated by spin vectors, bonding electrons by dots, the metallic orbitals by open circles):

		4*d*					5*s*	5*p*		
Sn A	$v = 4$	↑↓	↑↓	↑↓	↑↓	↑↓	•	•	•	•
Sn B	2	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	•	•	○
Sn C	0	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	○	○

Sn A has the maximum possible valency, 4. Because of their greater stability, it is essentially the *d* orbitals that are occupied by the five unshared electron pairs. The four bonding electrons occupy the four remaining orbitals, which may conveniently be considered to be hybridized to sp^3 tetrahedral bond orbitals. There is no extra orbital to serve as the metallic orbital, and so this quadrivalent tin would not, alone, form a metallic phase. Instead (Pauling 1947) it forms the essentially non-metallic variety of tin, grey tin, which has the diamond arrangement, each atom being surrounded tetrahedrally by four neighbours, to which it is attached by non-resonating single bonds.

The metallic form of tin, white tin, consists largely of the bivalent atoms Sn B, which have a metallic orbital. In white tin each atom has co-ordination number 6, and the bonds resonate among the alternative positions. It is the energy of this resonance which in the main compensates for the difference in bond energy between four bonds and two; some stabilization of bivalent tin also results from the fact that in it two electrons occupy essentially the 5*s* orbital, which has lower energy than the 5*p* orbitals.

White tin, to which the valency 2·44 has been assigned (Pauling 1938), does not contain bivalent Sn B alone, but also some Sn A. It seems likely that these two kinds of tin atoms occur in the ratio 3 : 1, leading to the average valency 2·5, a value in essential agreement with that suggested before.

Thus in general the non-integral metallic valencies shown by the transition elements are to be interpreted as averages, corresponding to resonance of each atom among two or more electronic structures with integral valencies. The relative weights of the different structures would be for a given metallic phase such as to minimize the energy of the system. A simple treatment involving consideration of the energy quantities involved, and especially of the resonance energy as affected by the ratio of metallic orbitals to atoms, permits the conclusion to be drawn that a 3 : 1 ratio of a lower-valent form with a metallic orbital and the higher-valent form without this orbital is reasonable; but deviations from this ratio are to be expected, and tin in its various alloys might show any valencies from 2 to 4. The zero-valent form Sn C has no virtues, and presumably does not make a contribution in any alloy of tin, but the analogous forms of certain other metals, with valency 4 less than the maximum, are important.

Similar valence structures A, B, etc., may be written for gallium, zinc, and copper and their congeners:

		3*d*					4*s*	4*p*		
Ga A	$v = 5$	⇅	⇅	⇅	⇅	•	•	•	•	•
Ga B	3	⇅	⇅	⇅	⇅	⇅	•	•	•	○
Ga C	1	⇅	⇅	⇅	⇅	⇅	⇅	•	○	○
Zn A	$v = 6$	⇅	⇅	⇅	•	•	•	•	•	•
Zn B	4	⇅	⇅	⇅	⇅	•	•	•	•	○
Zn C	2	⇅	⇅	⇅	⇅	⇅	•	•	○	○
Zn D	0	⇅	⇅	⇅	⇅	⇅	⇅	○	○	○
Cu A	$v = 7$	⇅	⇅	•	•	•	•	•	•	•
Cu B	5	⇅	⇅	⇅	•	•	•	•	•	○
Cu C	3	⇅	⇅	⇅	⇅	•	•	•	○	○
Cu D	1	⇅	⇅	⇅	⇅	⇅	•	○	○	○

These tend to be combined to produce effective metallic valencies of 3·5 for gallium, 4·5 for zinc, and 5·5 for copper (and their congeners), but other valencies are also shown by these versatile metals.

Pivoting valence-bond resonance

In ordinary covalent molecules and crystals the co-ordination number of an atom is equal to its covalency, because the resonance energy of synchronized resonance is not great enough to overcome the additional repulsive energy between non-bonded atoms. However, in an electron-deficient substance, with extra orbitals permitting uninhibited resonance, the resonance energy often becomes great enough to stabilize a structure with increased ligancy (co-ordination number)—the ligancy then usually increasing to a value greater than the number of orbitals.†

It is of interest to note that *in order for a given atom to increase its ligancy beyond its covalency it is not necessary that this atom have an extra orbital: it is instead sufficient for the atoms that surround it to have extra orbitals.* The valence bonds of the central atom may then resonate among their alternative positions by pivoting about the central atom.

It is this sort of pivoting resonance that permits the 25 % contribution of Sn A to white tin, in which each atom has ligancy 6. Other examples are AuSn (Pauling 1947), in which each tin atom is surrounded by a trigonal prism of gold atoms, and Mg_2Sn, with the fluorite structure (Pauling 1923), in which each tin atom is surrounded by a cube of magnesium atoms. The interatomic distances show that in each case the tin is quadrivalent, and its four bonds are permitted to resonate among the six or eight positions by use of the metallic orbitals of the gold or magnesium atoms. Resonance of this sort also characterizes many carbides, such as cementite, in which each carbon atom is surrounded by six iron atoms.

The stable atomic orbitals

We now interpret the ferromagnetic saturation moments as indicating that as many as seven hybrid orbitals have such a nature (concentration in bond directions, permitting large overlapping with orbitals of adjacent atoms) as to make them useful as bond orbitals (including the metallic orbital), and that the remaining two orbitals, essentially *d* in character, are available for occupancy by unshared electrons. These orbitals are of such a nature as to lead to positive rather than negative exchange integrals (small in magnitude, however), thus causing the parallel orientation of spins of adjacent atoms that leads to ferromagnetism.

Often, however, one or more additional *d* orbitals are used as atomic orbitals. This occurs as the result of the stabilizing influence of certain configurations of the atomic electrons and of certain molecular or crystal structures requiring only a smaller valency (smaller number of bond orbitals than seven). Thus *chromium*, with six outer electrons, might have any valency from 6 to 1, the non-valence electrons occupying the atomic orbitals. By Hund's rule of maximum multiplicity for a partially filled subgroup these $6-v$ atomic electrons would remain unpaired, and would require one 3*d* orbital apiece. If conditions were such as to favour a large valency the element would have its maximum valency 6. If they were to favour a small valency we might at first expect two electrons to occupy the two stable atomic orbitals, leading to $v = 4$. However, if three *d* orbitals combine with sp^3 to produce seven most satisfactory bond orbitals, only 2·14*d* orbitals would be required to produce five bond orbitals (for $v = 4$ and the metallic orbital) with the same hybrid character, or 1·71 to produce four. Hence by the change to a lower valency another *d* orbital is released for use as a stable atomic orbital, and a third electron might well occupy it. Indeed, the energy stabilizing the atomic state of maximum multiplicity is three times as great for the quartet configuration d^3 as for the triplet configuration

---

† V. Schomaker, private communication; also Conference on the Chemical Bond, Paris, 15 April 1948.

d^2, and hence there is good reason to expect the lower-valent state of chromium to have three atomic electrons and valency 3. Both sexivalent and tervalent chromium exist, as forms of the elementary metal, with the structures A2 and A3, respectively. In addition to the evidence of the interatomic distances (Pauling 1947), the value 3 for the valency of the low-valent state is supported by the value 3 magnetons found for the magnetic moment of chromium alloyed with gold† (so far as I know, the saturation moment of A3 chromium has not been reported).

Manganese by a similar argument would be expected to have the metallic valencies 6 and 4, as observed (Pauling 1947). The corresponding valencies 6 and 5 for *iron* do not occur as distinct modifications of the elementary substance, but instead both contribute, in ratio 78 : 22, to the average valency 5·78 shown for α-iron by the ferromagnetic saturation moment 2·22. It may be pointed out that the quinquevalent state of iron is not expected to be so stable as the lower-valent states of chromium and manganese because the decrease in valence by 1 would liberate only 0·43 of a *d* orbital, and the introduction of an additional atomic electron would hence weaken the bond orbitals by decreasing their *d* character.‡

Cobalt in its normal metallic form involves resonance between the two following sexivalent structures, in the ratio 35 : 65, as indicated by the saturation moment 1·71 magnetons:

		3d					4s	4p		
Co A	v = 6	↑	↑	↑	•	•	•	•	•	•
Co B	6	↑↓	↑	•	•	•	•	•	•	○

The contribution of Co A is presumably limited to 35 % by the destabilizing effects of absence of a metallic orbital and decreased *d* character of the bond orbitals, which oppose the stabilizing effect of the quartet atomic state. In the same way *nickel* involves resonance between the two following structures, in the ratio 30 : 70 (saturation moment 0·61):

		3d					4s	4p		
Ni A	v = 6	↑↓	↑	↑	•	•	•	•	•	•
Ni B	6	↑↓	↑↓	•	•	•	•	•	•	○

The decrease in the contribution of A from 35 % (for cobalt) to 30 % is presumably due in the main to the smaller amount of resonance stabilization for the triplet state than for the quartet state.

We have thus concluded from the foregoing arguments that the high-valent states of metallic chromium, manganese, cobalt, and nickel are all sexivalent states, with only iron having the slightly lower valency 5·78, in contradiction to the assumption previously made (Pauling 1938), in ignorance of a sound theory of the nature of metals, that the valency 5·78 applies to all the transition elements. The ferromagnetic data indicate that in alloys of iron and cobalt the average valency remains about 5·78 from 0 to 25 % cobalt (the moment increasing linearly from 2·22 to 2·48), and then increases to 6 at about 50 % Co (the moment remaining roughly constant at 2·48), and thereafter remains 6 (the moment decreasing rapidly). The detailed discussion of the electronic structure of these alloys in terms of Fe A, Fe B, Co A and Co B is complicated, and requires consideration also of the ions Fe^+ A, Fe^- A, etc.

---

† Personal communication from Professor P. Néel.

‡ The quinquevalent state with a pure *p* metallic orbital would have somewhat more *d* character in the bond orbitals.

THE STABILITY OF STRUCTURES IN RELATION TO BOND NUMBER

The factors that determine the choice of one from among alternative crystal structures of a metal or intermetallic compound have not been well understood. In the resonating-valence-bond theory of metals a very important energy quantity contributing to the stability of the systems is the energy of resonance of bonds among interatomic positions. It is clear that some modes of resonance would make larger contributions to the energy than others, and that in particular a simple ratio of number of bonds to number of positions would be exceptional. The simplest ratios, in order, are $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $\frac{1}{4}$, $\frac{3}{4}$, etc. We are thus led to the principle that *a special stability is associated with bond numbers* $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $\frac{1}{4}$, $\frac{3}{4}$, etc.†

For example, tin, with $v = 2·5$, crystallizes with a unique atomic arrangement, in which each atom has six ligates, four at 3·016 Å and two at 3·175 Å. These distances have been used (1947) in assigning the bond numbers 0·48 and 0·26 to these bonds. It is clear that these bond numbers can be taken as $\frac{1}{2}$ and $\frac{1}{4}$, and that the choice of the structure and the value of its axial ratio (which determines the relative lengths of the two kinds of bonds) are the result of the effort of the tin atom to use its valency 2·5 in the formation of stable bonds with simple fractional bond numbers.

Zinc and cadmium have the A 3 structure, but with abnormally large axial ratio (1·856 for zinc) instead of the value 1·633 corresponding to close packing. From the distances 2·660 Å (for six bonds) and 2·907 Å (for the other six) the bond numbers 0·54 and 0·21 have been deduced. If the axial ratio were normal each of the twelve bonds formed by a zinc atom with $v = 4·5$ would have bond number $\frac{3}{8}$. The assumption of the distorted structure permits a split into two classes with the more stable bond numbers $\frac{1}{2}$ and $\frac{1}{4}$ (or, probably, $\frac{1}{6}$, with the average valency of zinc equal to 4).

A similar explanation can be given of the occurrence of other unusual structures, shown by mercury, gallium, indium, β-tungsten, β-uranium, α-manganese, and β-manganese. There is, of course, a close relation between the principle of the stability of simple fractional bond numbers and the filling of the Brillouin polyhedra corresponding to strong perturbations of the wave functions for the resonating valence bonds, discussed below.

Because of the occasional occurrence of a difference in hybridization of the bond orbitals involved in bonds of different classes, leading to a difference in single-bond radius, the determination of the bond numbers from experimental values of the bond lengths involves some uncertainty.

A NEW SYSTEM OF METALLIC RADII

The interpretation of experimental values of interatomic distances for metals in terms of bond numbers, with use of the equation

$$R_n = R_1 - 0·300 \log_{10} n, \tag{9a}$$

---

† The importance of half bonds (with $n = \frac{1}{2}$) has been emphasized by Rundle (1947).

has permitted considerable progress to be made, but has been hampered by uncertainty as to whether or not a significant change in the single-bond metallic radius of an element accompanies a change in its valence. The system described here is designed to eliminate this difficulty by permitting the prediction of the metallic radius for any element in any valence state.

This system is based on the observations (Pauling 1947) that a linear relation between single-bond radius and atomic number holds for bonds of constant hybrid character, and that for an element the single-bond radius is (at least approximately) linearly dependent on the d character of the dsp hybrid bond orbitals.

The values of $R_1(sp^3)$, the single-bond radius for sp^3 bonds, for the iron-transition elements are given by the equation

$$R_1(sp^3) = 1{\cdot}825 - 0{\cdot}043z, \tag{10a}$$

in which $z = Z - 18$ is the number of outer electrons in the neutral atom (outside of the argon shell). This equation represents the straight line in figure 3 of Pauling (1947).

The value 1·165 Å for iron with valence 5·78 is an average for 78 % of sexivalent iron, with seven d^3sp^3 orbitals, and 22 % of quinquevalent iron, with d^2sp^3 orbitals (in each case including one metallic orbital); this value hence corresponds to 39·7 % d character, and with the assumed linear relation leads to 1·138 Å for d^3sp^3 orbitals (42·9 % d character). Similarly the value 1·162 Å for sexivalent cobalt with 39·5 % d character (average of 65 % d^3sp^3 and 35 % d^2sp^3, calculated from the magnetic moment) leads to 1·138 Å for the d^3sp^3 radius, and the value 1·137 Å is found for nickel from the sexivalent single-bond radius 1·154 Å for 40·0 % d character (70 % d^3sp^3 and 30 % d^2sp^3).

The essential equality of these three values suggests† that 1·138 Å may be taken as R_1 for d^3sp^3 orbitals for all the elements of the first long period:

$$R_1(d^3sp^3) = 1{\cdot}138. \tag{10b}$$

Equations (10a) and (10b) and the assumed linear relation between R_1 and the amount of d character (δ) lead to the equation

$$R_1(\delta, z) = 1{\cdot}825 - 0{\cdot}043z - (1{\cdot}600 - 0{\cdot}100z)\,\delta. \tag{10c}$$

The radii for pure p bonds are about 0·030 Å larger than for sp^3 bonds:

$$R_1(p) = R_1(sp^3) + 0{\cdot}030. \tag{10d}$$

---

† The treatment adopted here must be considered as an approximation, inasmuch as it involves neglect of factors that without doubt have some effect on the interatomic distances. Among these factors is the difference in the ability of single unshared electrons and electron pairs to compete with bonding electrons for the stable orbitals. It is not unlikely that the radius 1·165 for iron corresponds to an amount of d character larger than the assumed 39·7 %, because of the promotion of the atomic electrons to orbitals with less than 100 % d character. Moreover, it is likely, from the results of the analysis of the energy quantities for the alkali metals, that the metallic orbital has less d character than the bond orbitals, instead of the same amount as assumed. The contribution of the ionic aspects of the atoms, with one bond more or less than normal, has also not been considered explicitly. Incorporation of these effects in the treatment of interatomic distances would make the system much more complicated, without greatly changing the conclusions derived from its application.

Application of equation (10*c*) to the observed single-bond radii of *scandium, titanium* and *vanadium* (1·439, 1·324, 1·224 Å) leads to 20, 27 and 35 % of *d* character, respectively (table 5). The gradual increase presumably results from the increasing stability of the 3*d* orbitals relative to 4*s* and 4*p*.

TABLE 5. SINGLE-BOND RADII AND AMOUNT OF *d* CHARACTER FOR TRANSITION ELEMENTS

Sc	Ti	V	Cr	Mn	Fe	Co	Ni
1·439 (3)	1·324 (4)	1·224 (5)	1·176 (6, 3)	1·171 (6, 4)	1·165 (5·78)	1·162 (6)	1·154 (6)
20	27	35	39	40·1	39·7	39·5	40·0
Y	Zr	Cb	Mo	Tc	Ru	Rh	Pd
1·616 (3)	1·454 (4)	1·342 (5)	1·296 (6)	1·271 (6)	1·246 (6)	1·252 (6)	1·283 (6)
19	31	39	43	46	50	50	46
Lu	Hf	Ta	W	Re	Os	Ir	Pt
1·557 (3)	1·442 (4)	1·343 (5)	1·304 (6)	1·283 (6)	1·260 (6)	1·265 (6)	1·295 (6)
19	29	39	43	46	49	49	44

The observed single-bond radius for both sexivalent and tervalent *chromium* is 1·177 Å, which corresponds to 39·0 % *d* character. This might result for sexivalent chromium from a large (42 %) contribution of the set of orbitals d^2sp^3 (without a metallic orbital), or for a disproportion in the d^3sp^3 set, with the metallic orbital having more than the average amount of *d* character. In tervalent chromium three of the orbitals (presumably largely *d*) are occupied by the three atomic electrons, and the d^2sp^3 set remaining for the three bonds and the metallic orbital may produce the hybrid character that is most stable.

For *manganese* the quadrivalent single-bond radius 1·163 Å corresponds to 40·1 % *d* character. The sexivalent single-bond radius is the same, to within the accuracy (about 0·010 Å) of its determination from the interatomic distances in the crystals (β-manganese and α-manganese) in which sexivalent and quadrivalent manganese occur together.

The metals of the first ascending branch

If the assumption is made that the bond orbitals and one metallic orbital (except for the state with maximum valence, which has no metallic orbital) have the same hybrid character, values of the radii for the various pure valence states of the metals of the first ascending branch, from copper to germanium, can be calculated by use of equations (10*c*) and (10*d*). These values are given in table 6. There are also given the values interpolated for resonance between the state of maximum valency (with no metallic orbital) and the next state (with valency two less, and with a metallic orbital) in the ratio 25 : 75, the number of orbitals being included in the calculation as a weight factor.

For *copper* the predicted value of R_1 for the hybrid state with valency 5·5 is 1·172 Å. This is in nearly exact agreement with the observed value, $R_{(\mathrm{CN}12)} = 1{\cdot}276$ Å, which corresponds to $R_1 = 1{\cdot}174$ for $v = 5{\cdot}50$, and hence supports this value for the valency of copper in the elementary metal.

Zinc in the elementary metal has six ligates at 2·660 Å and six at 2·907 Å. If the strong bonds are assumed to have bond number $n = \frac{1}{2}$, the value of R_1 is 1·240 Å. This is close to the value 1·229 Å predicted for Zn B, with $v = 4$. If this quadrivalent zinc alone is represented, the weak bonds have $n = \frac{1}{6}$, and the value of R_1 calculated with the correction $-0{\cdot}300 \log \frac{1}{6} = 0{\cdot}234$ is 1·220 Å, again in good agreement with the predicted value. On the other hand, if there were 25 % contribution of Zn A, making $n = \frac{1}{4}$ for the long bonds, the value of R_1 would be 1·273 Å, which is much larger than the predicted radius for $v = 4{\cdot}5$, 1·214 Å. We hence conclude that elementary zinc is pure quadrivalent Zn B, and that the six strong and six weak bonds have bond numbers $\frac{1}{2}$ and $\frac{1}{6}$, respectively. The strong bonds formed by each atom involve resonance of three valencies among six positions, and the weak bonds involve resonance of one valency among six positions.

TABLE 6. METALLIC RADII OF ELEMENTS OF THE ASCENDING BRANCHES

	Cu	Zn	Ga	Ge
A	1·138 (7)	1·176 (6)	1·206 (5)	1·223 (4)
B	1·185 (5)	1·229 (4)	1·266 (3)	1·253 (2)
C	1·227 (3)	1·309 (2)	1·296 (1)	
D	1·352 (1)			
	1·172 (5½)	1·214 (4½)	1·248 (3½)	1·244 (2½)
	Ag	Cd	In	Sn
A	1·303 (7)	1·343 (6)	1·377 (5)	1·399 (4)
B	1·353 (5)	1·400 (4)	1·442 (3)	1·434 (2)
C	1·396 (3)	1·485 (2)	1·477 (1)	
D	1·528 (1)			
	1·340 (5½)	1·384 (4½)	1·423 (3½) 1·456 (2½)	1·423 (2½)
	Au	Hg	Tl	Pb
A	1·303 (7)	1·345 (6)	1·387 (5)	1·430 (4)
B	1·351 (5)	1·403 (4)	1·460 (3)	1·540 (2)
C	1·393 (3)	1·490 (2)	1·540 (1)	
D	1·520 (1)			
	1·338 (5½)	1·388 (4½) 1·418 (3½)	1·419 (4) 1·439 (3½) 1·477 (2½) 1·497 (2)	1·506 (2½)

In the complex structure of *gallium* each atom has seven neighbours, at four different distances, 2·437 (1), 2·706 (2), 2·736 (2) and 2·795 Å (2). These were previously interpreted, with the assumed valency 3·44, as corresponding to bond numbers 1·21, 0·43, 0·38 and 0·31, respectively, leading to the single-bond radius 1·245 Å, and the value 1·251 Å would have been obtained if the valency had been taken as 3·50. The close agreement of this value with the predicted value 1·248 for valency 3·50 (table 6) thus supports this valency for elementary gallium.

If it is assumed that the valency 3½ is divided among one single bond, four ½-bonds, and two ¼-bonds the corresponding values of R_1 are 1·219, 1·263, 1·278 and 1·217 Å. The first and last values correspond to orbitals with 16 % d character, and the others to sp^3 orbitals. This suggests that the 0·25 d orbital provided by the 25 % contribution of Ga A is divided among the $1\frac{1}{2}$ orbitals of the single bond and two $\frac{1}{4}$-bonds, giving 16·7 % d character to them.

The observed interatomic distance 2·446 Å for *germanium*, with the diamond structure, was used as a fixed point in formulating equation (9a), the valency 4 being assumed. Germanium seems to be quadrivalent in all of its intermetallic compounds. A representative one is Mg_2Ge, with the Mg_2Sn (fluorite) structure. The germanium atom is surrounded by eight magnesium atoms, and its four valences undergo pivoting resonance among the eight positions, leading to half bonds.

The radii of the heavier transition metals

For the palladium-transition metals we write the equations

$$R_1(sp^3) = 2{\cdot}001 - 0{\cdot}043z, \tag{11a}$$

$$R_1(d^3sp^3) = 1{\cdot}303, \tag{11b}$$

$$R_1(\delta, z) = 2{\cdot}001 - 0{\cdot}043z - (1{\cdot}627 - 0{\cdot}100z)\,\delta, \tag{11c}$$

$$R_1(p) = R_1(sp^3) + 0{\cdot}035. \tag{11d}$$

The corresponding equations for the platinum-transition metals are

$$R_1(sp^3) = 1{\cdot}850 - 0{\cdot}030z, \tag{12a}$$

$$R_1(d^3sp^3) = 1{\cdot}303, \tag{12b}$$

$$R_1(\delta, z) = 1{\cdot}850 - 0{\cdot}030z - (1{\cdot}276 - 0{\cdot}070z)\,\delta, \tag{12c}$$

$$R_1(p) = R_1(sp^3) + 0{\cdot}110. \tag{12d}$$

Equations (11a) and (12a) are based upon the values 1·399 and 1·430 for $R_1(sp^3)$ for tin and lead, respectively, with slopes as indicated by the tetrahedral radii for the heavier elements. The value 1·430 for quadrivalent lead was obtained by adding to the radius for quadrivalent tin the difference 0·031 Å between the Mg-Pb distance in Mg_2Pb and the Mg-Sn distance in Mg_2Sn. The simple form of equations (11b) and (12b) was assumed in analogy with (10b), and the equal values of the constant were obtained from the observed distances for silver and gold, both assumed to have valence $5\frac{1}{2}$.

Revised values† of the single-bond radii for the elements molybdenum to palladium and tungsten to platinum are given in table 5. These differ from those previously published (Pauling 1947) in that the valency is assumed‡ to be 6 instead of 5·78. The course of the radii for these sequences differs from that in the sequence chromium to nickel in that the contraction shown in the preceding sequences continues, a minimum being reached at ruthenium and osmium. It is likely that this further contraction beyond molybdenum and tungsten is due to the appropriation of an amount of d character greater than 42·9 % (for d^3sp^3 orbitals) by the bond

---

† The value for technetium is from Mooney (1948):

$$\text{Tc-Tc} = 2{\cdot}706\,(6),\ 2{\cdot}738\,(6);\ R_1(\text{CN}\,12) = 1{\cdot}361\,\text{Å}.$$

‡ Possibly slightly different valencies should be assumed, especially for ruthenium and osmium, and a small change may be required when reliable magnetic data become available.

orbitals, probably at the expense of both the atomic orbitals (occupied by unpaired electrons) and the metallic orbital. The effect decreases at palladium and platinum, with atomic orbitals mainly occupied by pairs of electrons, which tend to cause them to retain 100 % *d* character. It is seen from table 5 that the amounts of *d* character calculated by application of equations (11*c*) and (12*c*) show a reasonable trend, and are almost exactly the same for the two series of elements. The *d* character is equal to that for d^3sp^3 bonds at molybdenum and tungsten, and reaches its maximum of 50 % at ruthenium, rhodium, osmium, and iridium.

The metals of the silver and gold ascending branches

Values of the single-bond radii for the metals of the series silver to tin and gold to lead are also given in table 6, as calculated by equations (10*c*) and (11*c*) for the pure valence states A, B, C and D and for certain intermediate valencies, corresponding to resonance among these states.

The values for *silver* and *gold* with $v = 5\frac{1}{2}$ agree closely with the experimental values, 1·340 and 1·337 Å, respectively, upon which the equations were in part based.

For *cadmium* the observed interatomic distances 2·973 (6) and 3·287 Å (6) indicate bond numbers $\frac{1}{2}$ and $\frac{1}{6}$, respectively, as for zinc, and lead to the values 1·397 and 1·410 Å for R_1, in good agreement with the value 1·400 Å given in the table for $v = 4$. Bond number $\frac{1}{4}$ for the longer bonds would lead to $R_1 = 1{\cdot}463$, in disagreement with the value 1·384 for $v = 4\frac{1}{2}$. Hence we conclude that in the elementary metal cadmium, like zinc, is quadrivalent.

Mercury, with interatomic distances 2·999 (6) and 3·463 (6), appears to have valency $3\frac{1}{2}$. With bond numbers $\frac{1}{2}$ and $\frac{1}{6}$, respectively, these distances lead to $R_1 = 1{\cdot}410$ and 1·498, the latter being much too large for $v = 4$ ($R_1 = 1{\cdot}403$), whereas bond numbers $\frac{1}{2}$ and $\frac{1}{12}$ lead to $R_1 = 1{\cdot}410$ and 1·408, in approximate agreement with the value 1·418 for $v = 3\frac{1}{2}$. The decrease in valency from cadmium to mercury conforms to a general trend toward smaller metallic valencies with increasing atomic number in a group of elements.

In *indium* each atom has four ligates at 3·242 and eight at 3·370 Å. The first of these distances leads with $n = \frac{1}{4}$ to $R_1 = 1{\cdot}440$, a reasonable value in comparison with $R_1 = 1{\cdot}442$ for $v = 3$ (table 6). However, the eight longer bonds must have $n < \frac{1}{4}$, so that the valency is less than 3. With $n = \frac{3}{16}$ for these bonds their value of R_1 is 1·446, the weighted average of this value and that for the shorter bonds, 1·456, agreeing exactly with the value for $v = 2\frac{1}{2}$ in table 6; the valency of indium in the elementary metal is hence indicated to be $2\frac{1}{2}$.

In grey *tin* the distances 3·016 (4) and 3·175 (2) lead with $n = \frac{1}{2}$ and $\frac{1}{4}$ to $R_1 = 1{\cdot}418$ and 1·407, respectively, in approximate agreement with the interpolated value 1·423 for $v = 2\frac{1}{2}$. The distances 2·980 (2) and 3·198 (6) for γ-tin (containing a small amount of mercury), with $n = \frac{1}{2}$ and $\frac{1}{4}$, similarly give $R_1 = 1{\cdot}400$ and 1·418.

Thallium, with R (CN 12) = 1·712 Å, seems to be approximately bivalent. With $n = \frac{1}{6}$ the value 1·478 is found for R_1, slightly smaller than that in table 6 for $v = 2$, 1·497. The value R (CN 12) = 1·746 Å for *lead* gives $R_1 = 1{\cdot}512$ for $v = 2$, and $R_1 = 1{\cdot}542$ for $v = 2\frac{1}{2}$; comparison with the values in table 6 suggests that the valency lies between these limits. The conclusions for thallium and lead are uncertain because of uncertainty in the radii for pure p bonds. The value of the constant in equation (11d) is indicated not only by the single-bond radius for bismuth but also by the distances in the diatomic hydrides (M-H distance minus $0{\cdot}300 = 1{\cdot}570$ for TlH, 1·539 for PbH, 1·509 for BiH; table III of Pauling 1947). By decreasing these radii by 0·03 Å exact agreement would be obtained for both thallium and lead with valency 2.

INTERATOMIC DISTANCES IN INTERMETALLIC COMPOUNDS

The discussion of interatomic distances is less simple for intermetallic compounds than for pure metals; among the complicating factors are the partial ionic character of bonds, the transfer of electrons, with consequent changes in valency, and the preferential use of the valencies of an atom in the formation of strong bonds rather than weaker ones. These factors, which of course participate in minimizing the energy of the system, usually operate to decrease the interatomic distances. Their effects may be illustrated by some examples.

In a discussion of cementite, Fe_3C, it was pointed out (Pauling 1947) that the observed interatomic distances, Fe-C = 2·01 Å and Fe-Fe = 2·62 Å (average), differ from those predicted from the valencies and radii of iron and carbon, Fe-C = 2·04 Å and Fe-Fe = 2·59 Å, by ± 0·03 Å, and that the lack of agreement might indicate that the structure is under some strain, compressing the Fe-C bonds and stretching the Fe-Fe bonds. An alternative and in some ways preferable explanation is that, in order to form as strong bonds as possible with the carbon atoms, the two corresponding orbitals of the iron atoms undergo a change in hybridization with increase in the amount of d character and decrease of the iron radius by 0·03 Å, and that a compensating decrease in the amount of d character (by one-half as much) occurs in the four orbitals involved in forming Fe-Fe bonds, accompanied by an increase in radius by 0·015 Å. The same explanation has been advanced (Pauling & Soldate 1948) for the observed shortening of the Fe-Si distances in the cubic crystal FeSi; and in addition it has been suggested that in this crystal the valency of silicon is used entirely in forming bonds with iron, the Si-Si separation of 2·76 Å, which indicates bond number 0·20, being assumed rather to be a non-bonded contact required by the geometry of the atomic arrangement.

The substance FeB, which may serve as another example, forms orthorhombic crystals (Bjurström 1933) in which each boron atom has two boron ligates at 1·77 Å and six iron ligates at 2·15 ± 0·03 Å. These distances correspond to half bonds, for which the calculated distances are 1·78 and 2·15 Å, respectively, and the covalency of boron is hence indicated to be 4. We might thus be led to assume that, in order to allow the formation of the maximum number of strong bonds (involving boron) permitted by the number of bonding orbitals of the boron atom, an electron is transferred from iron to boron. An alternative explanation is that the B-Fe bonds have bond number $\frac{1}{3}$, the iron radius being decreased by 0·11 Å by an increase in d character of the two iron orbitals involved in these bonds, and the boron atom remaining tervalent. Such a large decrease in the iron radius seems unlikely, however, and it is instead probable that both effects operate, there being a smaller change in hybridization and also some electron transfer in this crystal.

Each iron atom in FeB has six boron ligates, six iron ligates at 2·64 Å, and four iron ligates at 2·94 Å, corresponding to valency about 5·5.

The interpretation of simple structures of intermetallic compounds is made difficult by the metrical relations between non-equivalent interatomic distances that prevent them from independently assuming the values corresponding to their bond type. The B 32 structure, shown by LiGa and several other compounds of elements of groups I and III *b* of the periodic table, consists of two interpenetrating diamond arrangements: each atom is surrounded by four atoms of the same kind and four of the other kind, at the tetrahedral corners of a cube, and there is thus only one bond distance (2·68 Å in LiGa) for three kinds of bonds (Li-Li, Li-Ga and Ga-Ga), which hence are expected to be strained. Nevertheless the bond number calculated from this distance for the strongest bonds (Ga-Ga) should be approximately correct.

Several investigators have suggested that in the B 32 substances the alkali metal loses its valence electron to the III *b* metalloid, which then acts as a fourth-group element, and forms four single bonds with its four similar neighbours. However, this structure would lead to interatomic distances much smaller than those observed—to 2·53 Å for LiGa, for example, instead of 2·68 Å. The corresponding Ga-Ga bond number, 0·57, calculated by equation (9), suggests instead that the bonds are half bonds, or perhaps $\frac{5}{8}$-bonds. Presumably the lithium atom uses its valence electron in forming bonds with gallium atoms (four Li-Ga $\frac{1}{4}$-bonds), the Li-Li contacts being essentially non-bonded. The valency of gallium is 3 (for Ga-Ga half bonds) or $3\frac{1}{2}$ (for $\frac{5}{8}$-bonds).

The compound PtTl (Zintl & Harder 1935) illustrates the phenomenon of the increase in valency of an element (thallium) in response to a change in its environment. This substance forms hexagonal crystals, the unit containing 3 Pt, 1 Tl I, and 2 Tl II, with ligation as follows: Pt, 4 Pt at 2·803 Å, 2 Tl I at 2·803 Å, and 4 Tl II at 2·83 Å; Tl I, 6 Pt at 2·803 Å; Tl II, 6 Pt at 2·83 Å. The Pt-Pt distance corresponds to $n = 0·44$, calculated by equation (9) with $R_1 = 1·295$. Presumably platinum has its normal valency 6, which is used in forming Pt-Tl bonds except for the amount used in the Pt-Pt bonds. The bond number $n = 0·44$ would lead to 4·24 as the average valency of thallium in this compound; it is likely, however, that the Pt-Pt bonds are half bonds, and that the valency of thallium is 4—twice that for the elementary substance. The predicted value of R_1 for quadrivalent thallium is 1·419 Å, and the calculated bond numbers for Pt-Tl I and Pt-Tl II are 0·71 and 0·64, respectively. The small differences between these values and the value $\frac{2}{3}$ corresponding to the assumed valencies and between 0·44 and $\frac{1}{2}$ for the Pt-Pt bonds may be attributed to a change in hybridization of the orbitals of platinum, in response partially to the requirement of the structure that the Pt-Pt and Pt-Tl I distances be equal and partially to the effort to use better bond orbitals (with more *d* character) for the stronger Pt-Pt bonds than for the weaker Pt-Tl bonds.

There are many interesting problems involving solid solutions that might be discussed by use of the new system of radii. Of these I select one pointed out to me by Mr E. C. Ellwood, of King's College, Newcastle-upon-Tyne. The solid solution of zinc in aluminium, with the Al structure, extends from 0 to 66 atomic % zinc. On cooling below 365° C the phase containing approximately 40 at. % zinc separates into two phases, both of which are solid solutions with the Al structure; at 275° C one phase (α) contains only 17 at. % zinc, and the other (α') contains 59 at. % zinc. The values of the lattice constants indicate that the effective radius of zinc is smaller in the α' phase than in the α phase.

An explanation that may be suggested of these facts is that solid solutions of a quadrivalent metal (zinc) in a tervalent metal (aluminium) tend to be unstable because of the difficulty of saturating the valency of isolated quadrivalent atoms by bonds to its lower-valent ligates. With zinc as the solute an increase in free energy at the lower temperatures would accompany the separation into the zinc-poor α phase, in which the versatile zinc atoms tend to assume the valency 3 (less stable, however, for them than their normal valency) in order to fit into the aluminium structure, and the zinc-rich α' phase, in which the concentration of zinc atoms is great enough to permit the extra valency of zinc to be satisfied through the formation of Zn-Zn bonds.

BRILLOUIN ZONES AND THE RESONATING-VALENCE-BOND THEORY

In the customary treatment of metals it is shown that the interaction of the metallic electrons with the atomic nuclei (and their strongly bound electrons) arranged in a crystallographic array leads to perturbations of the energy that can be described as the concentration of the one-electron energy levels into Brillouin zones. It is interesting that the same concentration into Brillouin zones results from the resonating-valence-bond theory, and, moreover, the successive zones (Brillouin polyhedra) contain the same numbers of electrons as given by the one-electron molecular-orbital theory.

Instead of formulating the wave function for a crystal as a sum of functions describing various ways of distributing the electron-pair bonds among the interatomic positions, as was done in the first section of this paper, let us formulate it in terms of two-electron functions describing a single resonating valence bond. A bond between two adjacent atoms a_i and a_j may be described by a function $\phi_{ij}(1,2)$ in which 1 and 2 represent two electrons and the function ϕ_{ij} may have the simple Heitler-London form

$$\phi_{ij}(1,2) = C\{\phi_i(1)\phi_j(2)+\phi_j(1)\phi_i(2)\}\{\alpha(1)\beta(2)-\beta(1)\alpha(2)\}, \qquad (13)$$

with ϕ_i and ϕ_j the appropriately hybridized bonding orbitals of atoms a_i and a_j and α and β the one-electron spin functions, or it may include ionic terms or have a still more complex form. The ϕ_{ij} could then be used in building antisymmetrized functions corresponding to the various possible valence-bond structures for the crystal. This procedure would lead to essentially the same results as those given by equations (1) and (2). However, a convenient approximation is made by neglecting the resonance energy of the bonding pairs of electrons with one another, and considering only the bond energy and the resonance energy of each bond among its alternative positions, without correlating its resonance with that of other bonds. The wave function for the crystal (not antisymmetrized) may then be written as

$$\psi(1, 2, \ldots, 2N) = \Phi_1(1, 2)\,\Phi_2(3, 4), \tag{14}$$

in which each of the functions Φ_k is a two-electron function representing a valence bond resonating through the crystal. The total energy, in this approximation, is the sum of the resonating-bond energies:

$$W = \Sigma W_k. \tag{15}$$

The form of the functions Φ_k may be closely similar to that of the molecular orbitals used in the simple theory of metals. If there are M interatomic positions in the crystal which might be occupied by any one of the N electron-pair bonds, then the M functions ϕ_{ij} may be combined into the M mutually orthogonal linear aggregates that approximate the solutions of the wave equation with inclusion of the interaction terms representing resonance. This combination can be effected with use of Bloch factors:

$$\Phi_k(1, 2) = C\phi_{\mathbf{r}}(1, 2)\, \mathrm{e}^{2\pi i \mathbf{P}_k \cdot \mathbf{r}/h} \tag{16}$$

Here $\mathbf{r}$ is the radius vector from the origin to a point R in the crystal, $\phi_{\mathbf{r}}$ is the electron-pair-bond function in the region near R, $\mathbf{P}_k$ is the momentum vector corresponding to the three quantum numbers k (the density of states being calculated in the usual way), h is Planck's constant, and C is the normalizing factor.

When the momentum vector $\mathbf{P}_k$ has such magnitude and orientation as to correspond to Bragg reflexion from an important crystallographic plane (with large structure factor for slow electrons) strong interaction between the resonating bond and the rest of the crystal will occur, leading to the displacement of energy levels and the formation of Brillouin zones. The striking agreement between the number of bonding electrons given by the new system of metallic valencies and the number of electrons contained in an important Brillouin polyhedron for the γ-alloys, α-manganese, β-manganese, and some other metallic substances with properties indicating filled zones (brittleness, low electrical conductivity) has been discussed in an earlier paper (Pauling & Ewing 1948).

REFERENCES

Bjurström, T. 1933 *Ark. Kemi Min. Geol.* **11**, A, no. 5; *Strukturbericht*, **3**, 12, 619.
Mooney, R. C. L. 1948 *Acta Crystal.* **1**, 161.
Pauling, L. 1923 *J. Amer. Chem. Soc.* **45**, 2777.
Pauling, L. 1931 *J. Amer. Chem. Soc.* **53**, 1367.
Pauling, L. 1938 *Phys. Rev.* **54**, 899.
Pauling, L. 1947 *J. Amer. Chem. Soc.* **69**, 542.
Pauling, L. 1948*a* *Nature*, **161**, 1019.
Pauling, L. 1948*b* *Victor Henri Memorial Volume.* Liège: Maison Desser.
Pauling, L. & Ewing, F. J. 1948 *Rev. Mod. Phys.* **20**, 112.
Pauling, L. & Sherman, J. 1937 *J. Amer. Chem. Soc.* **59**, 1450.
Pauling, L. & Soldate, A. M. 1948 *Acta Crystal.* **1**, 212.
Pauling, L. & Wilson, E. B. (Jr.) 1935 *Introduction to quantum mechanics*, chap. XIII. New York: McGraw Hill Book Co.
Rundle, R. E. 1947 *J. Amer. Chem. Soc.* **69**, 1327, 1719.
Slater, J. C. 1931 *Phys. Rev.* **38**, 1109.
Zintl, A. & Harder, A. 1935 *Z. Elektrochem.* **41**, 767; *Strukturbericht*, **3**, 618.

The polymerization of acetaldehyde at low temperatures

By J. C. Bevington and R. G. W. Norrish, F.R.S.

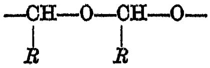
Department of Physical Chemistry, University of Cambridge

(*Received* 21 *August* 1948)

An account is presented of a study of the production of a high polymer during the solidification of acetaldehyde. Details are given of an experimental technique which leads to reproducible results. It is shown that the kinetics of the reaction indicate that the active centres have long lives. Although various substances are strong inhibitors of the polymerization, there is no evidence that a catalyst is needed to initiate the reaction. Traces of acetic acid have profound effects upon the properties of the polymer, and it is considered that the acid causes multiplication of centres by inducing rupture of carbon-oxygen bonds in the polymer. The polymer is unstable at room temperature and the course of depolymerization has been followed using both the solid polymer and solutions of it. The depolymerization is catalyzed by acetic acid, and it is suggested that the differences between the stabilities of various samples of polymer are to be explained by differences in their acetic acid contents. Possible mechanisms for the polymerization are discussed.

Introduction

Although high polymers of formaldehyde and cyclic trimers of certain other aldehydes have been known for many years, the first reports of high polymers of aldehydes other than formaldehyde were issued comparatively recently. These aldehydic polymers are substituted polyoxymethylenes of formula



$$-\underset{R}{\underset{|}{\mathrm{CH}}}-\mathrm{O}-\underset{R}{\underset{|}{\mathrm{CH}}}-\mathrm{O}-$$

derived by the opening of the carbon-oxygen double bond of the compound R.CHO. It was shown (Bridgman & Conant 1929) that *n*-butyraldehyde ($R = C_4H_9$) and certain other aldehydes were converted to high polymers when subjected to pressures of several thousand atmospheres, and that the materials gradually reverted to their monomers on releasing the applied pressure. The high polymer of acetaldehyde was discovered by Travers (1936) and Letort (1936) when purifying the aldehyde *in vacuo*. It was found that when the vapour was condensed on a surface at a temperature below the melting-point of the aldehyde (−123·5° C) or when the liquid was frozen, part of the aldehyde became polymerized; the aldehyde melted to a viscous liquid from which a solid polymer could be separated. The polymer slowly changed back to the monomeric aldehyde on standing at room temperature.

Since the first descriptions of the polymer appeared, further accounts of the reaction have been given (Letort & Duval 1943 *a*, *b* and Letort, Duval & Rollin 1947). It appears that the French workers performed only experiments of a qualitative nature, the degree of polymerization in a sample of aldehyde being estimated from the appearance. Certain of their results are conflicting; for example a statement that occurrence of polymerization depends upon the aldehyde being very pure, was later revised and it was said that the reaction is not affected by the source or the purity of the aldehyde, and is independent qualitatively of all catalytic influences. It was shown that polymerization does not occur if the aldehyde is at all times at temperatures above its melting-point, and that polymerization probably occurs while the aldehyde is being frozen.

Rigby, Danby & Hinshelwood (1948) attempted a more detailed study of the reaction. They cooled samples of liquid acetaldehyde to various temperatures and afterwards estimated the extent of polymerization by determining the viscosities of the samples under standard conditions; great irreproducibility was observed. It seems that no attempt was made to control the conditions very precisely, in particular no mention is made of excluding air from the aldehyde undergoing treatment. The marked inhibiting effects of certain substances were described, and the observation that solidification is essential for polymerization, was confirmed.

Staudinger (1936) first suggested that the high polymer of acetaldehyde is a substituted polyoxymethylene; this structure has been confirmed by consideration of the physical properties of the polymer, and further evidence for it has been provided by X-ray studies (Powell 1948) and by examination of the infra-red spectrum (Sutherland, Philpotts & Twigg 1946).

The object of the present work was to devise a technique which would afford accurate and reproducible results, and to apply it to a study of the kinetics of the polymerization.

EXPERIMENTAL

The apparatus and the method of working

The apparatus is shown in figure 1; the pumping unit was of conventional design and is omitted from the diagram. The essential feature of the apparatus is that purified liquid acetaldehyde could be stored *in vacuo* in the tube *C*, and that measured amounts of it could be introduced into the tube *F* by filling the bulb *D* with vapour to a measured pressure and then condensing the vapour in *F* cooled to −78° C. The liquid in *F* was then allowed to warm to 0° C. Meanwhile the tube *G* was evacu-

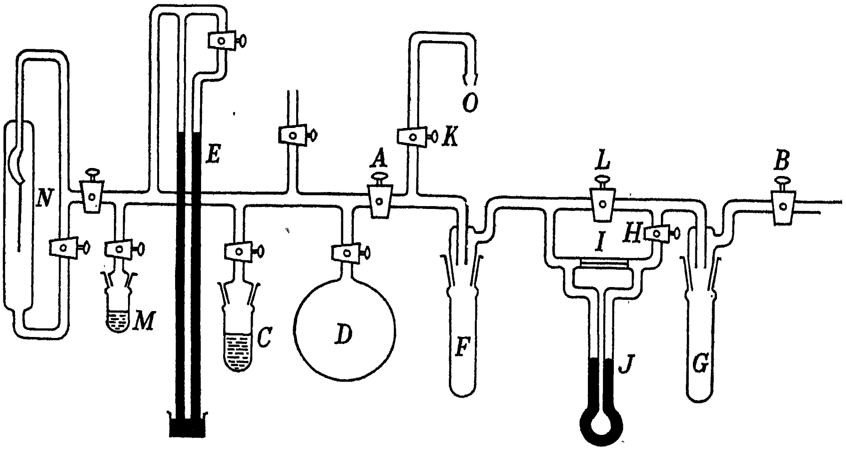


FIGURE 1. Apparatus.

ated to a pressure of 3×10^{-5} mm. Hg and flamed to rid it of adsorbed materials; this precaution was essential for reproducible results to be obtained. Tube *G* was then cooled in liquid nitrogen and aldehyde was distilled into it from tube *F* through the capillary *I*, the rate being kept constant by adjusting tap *H* to give a constant difference in the levels of mercury in *J*. The distillate was afterwards allowed to melt, and any polymer in it was precipitated by water; the polymer was dried *in vacuo* and weighed.

When volatile impurities were to be added to the aldehyde, bulb *D* was filled with the vapour from the reservoir *M*, the pressure being measured by the glass spoon gauge *N*. The vapour was condensed in *F* which was cooled in liquid nitrogen, this cooling agent could not be used for the subsequent condensation of the aldehyde because polymerization would have occurred during the solidification, and the monomer was condensed at −78° C as before.

Estimation of the molecular weights of the products

The molecular weights of certain of the products were compared by the viscosimetric method. For this purpose 0·4 % weight-to-volume solutions in *n*-butyl acetate were made, the solvent having been twice redistilled. The specific viscosities were determined at 20° C using an Ostwald viscometer. No assumptions were made concerning the exact form of the relationship between specific viscosity and average molecular weight.

Purification of materials

Acetaldehyde was twice fractionated through a long-column packed with Fenske rings. The final distillate was transferred to tube *C* (figure 1); approximately the first fifth was distilled away to the traps guarding the vacuum pumps, and in so doing dissolved air was removed. The tube was then connected to a similar one cooled to −78° C and the aldehyde was distilled into it; the last fifth was rejected. Tube *C* was cleaned and evacuated, and then cooled to −78° C; the aldehyde was distilled back to it, the first and last fractions being rejected as before. It was not possible to prepare batches of acetaldehyde with exactly the same reactivity, but the behaviour of separate batches was nearly the same and could be related by performing a standardizing experiment.

When foreign substances were required, the purest available material was chosen, and was further purified by vacuum distillation on the apparatus. The liquids were stored in tubes such as *M* (figure 1).

In early experiments, acetaldehyde was distilled directly into tube *F* through a fractionation column attached to the ground-joint *O*. This procedure was unreliable because (i) the reaction is very considerably affected by traces of impurities, and there is no guarantee of the column producing samples of aldehyde of identical purity in successive distillations; (ii) the aldehyde collected in *F* contains air (or nitrogen if the distillation is carried out in a stream of that gas).

The air could be removed by freezing the aldehyde and evacuating the tube, but it was found that satisfactory results could not be obtained. The technique was abandoned in favour of the one involving the use of a reservoir of liquid acetaldehyde and the measuring bulb.

EXPERIMENTAL RESULTS

The effects of foreign substances on the polymerization

Mixtures of acetaldehyde and volatile impurities were distilled from 0° C to liquid nitrogen. In each case 1·6 g. of aldehyde were distilled at the rate of 0·080 g./min. The results are presented in table 1; in this and other tables of results the experiments are arranged in the order in which they were performed. In the case of the run with 0·37 mole % acetone, the specific viscosity of a solution of the polymer was 1·13 compared with a value of 1·98 for the material produced in the succeeding blank experiment. It is evident that acetone reduced the average molecular weight of the product.

TABLE 1

added material	mole % added material	% yield of polymer
—	—	22·5
water	1·57	0
water	0·52	2·5
water	1·05	1·2
—	—	21·9
ethyl alcohol	1·05	0
ethyl alcohol	0·52	very small
—	—	23·7
acetone	1·05	1·2
acetone	0·52	6·2
acetone	0·37	11·3
—	—	23·7
paraldehyde	1·05	6·2
—	—	21·3

Ethyl alcohol also exerted its inhibiting effect when it was distilled into the receiver at the same time as the aldehyde but from a separate vessel. This suggests that the reduction in yield is due to some action of the inhibitor on the growing centre. In one experiment, pure acetaldehyde was distilled into the receiver cooled in liquid nitrogen and then a layer of alcohol was condensed on top of it; there was no reduction in the yield of polymer.

A study was made of the effect of adding acetic acid to the aldehyde prior to the polymerizing distillation. Table 2 shows that traces of acid increased the yield of polymer; in each case 1·6 g. of aldehyde were distilled at the rate of 0·080 g./min.

TABLE 2

mole % acetic acid	% yield of polymer
0	18·1
0·52	27·5
1·05	38·1
0	17·3
0·26	26·9
1·57	39·0
0	21·3

Acetic acid has a striking effect upon the physical properties of the polymer. Polymers made from aldehyde known to contain acetic acid were softer and less rubbery than those prepared from pure aldehyde; they dissolved more quickly in various solvents to give solutions less viscous than those made from polymers derived from pure aldehyde. It is clear that in increasing the yield of polymer, acetic acid reduced its average molecular weight considerably. Polymers prepared from monomer known to contain acetic acid were less stable than those from aldehyde substantially free from acid.

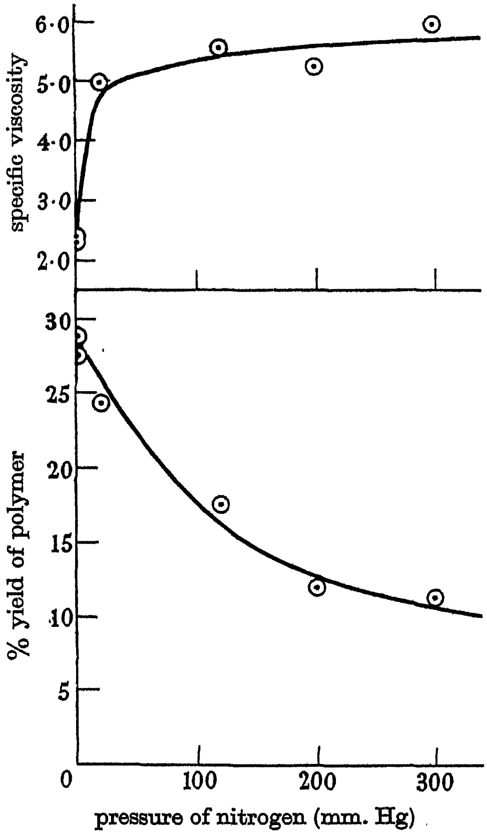


FIGURE 2. The effect of nitrogen on the yield and molecular weight of the polymer.

In one set of experiments, pure nitrogen was admitted to the tube *F* after the aldehyde had been condensed in it, and distillation to the tube *G* was performed in the presence of the added gas. The pressure of nitrogen was measured on the manometer *E*, and in each case 1·6 g. of aldehyde were distilled in 20 min. The results are shown in figure 2, and it can be seen that the molecular weight of the polymer (as indicated by the specific viscosity of the solution) was increased considerably by quite small amounts of added gas, although further increases in the quantity of nitrogen affected the molecular weight only slightly. The yield of polymer was depressed by the addition of nitrogen. When dry air was used instead of nitrogen, similar effects were observe . If, however, the aldehyde and air were allowed to stand at 0° C for 16 hr. before being distilled, the polymer had low molecular weight (for 200 mm. of air, the specific viscosity of the solution in *n*-butyl acetate was 1·17). This effect could be attributed to acetic acid formed by the oxidation of the liquid aldehyde.

The effects of varying the conditions during distillation

The weight of polymer produced was determined as a function of the rate of distillation of the acetaldehyde, and of the total weight of monomer used. In all cases the tube *F* (figure 1) was kept at 0° C and the tube *G* was cooled in liquid nitrogen. The results are shown graphically in figure 3; the experiment in which 1·6 g. of aldehyde distilled at 0·08 g./min., was treated as a blank, and was performed at the beginning and at the end of the series. The percentage of the monomer which polymerized, was low when small weights were used, but after the initial period there was a linear relationship between the weight of polymer produced and the time for which distillation was carried out.

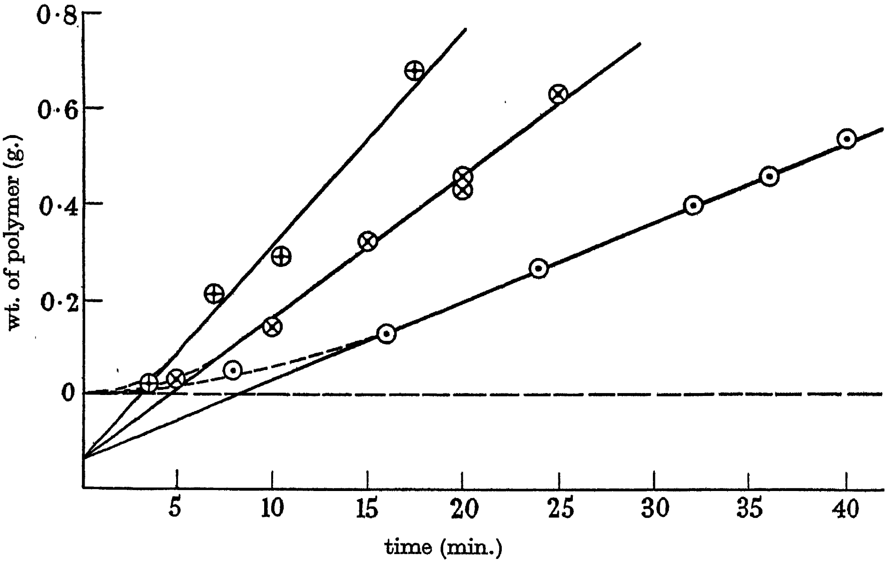


FIGURE 3. The weight of polymer produced as a function of the rate and of the time of distillation. ⊙ rate of distillation = 0·050 g./min.; ⊗ rate of distillation = 0·080 g./min.; ⊕ rate of distillation = 0·114 g./min.

A curve of the same form was obtained when acetaldehyde vapour was condensed directly from the bulb into the tube cooled in liquid nitrogen. The rate of condensation was fixed at about 0·080 g./min., by adjusting the taps so that the mercury level in the manometer rose at a constant rate. Figure 4 shows that there was little difference between these experiments and those in which the aldehyde was distilled into the receiver at the same rate.

Using the distillation technique, the dependence of the average molecular weight of the polymer upon the weight of acetaldehyde used was investigated. The specific viscosities of 0·4 % solutions of the polymer in *n*-butyl acetate were measured at 20° C. The results are shown in figure 5; it can be seen that the specific viscosity and therefore the average molecular weight of the polymer increased with the weight of aldehyde used, but tended to limiting values. In table 3 are shown the specific viscosities of solutions of polymers prepared by distilling 1·6 g. of acetaldehyde from 0° C to liquid nitrogen at various rates. They indicate that the limiting value of the average molecular weight of the polymer was independent of the rate of distillation.

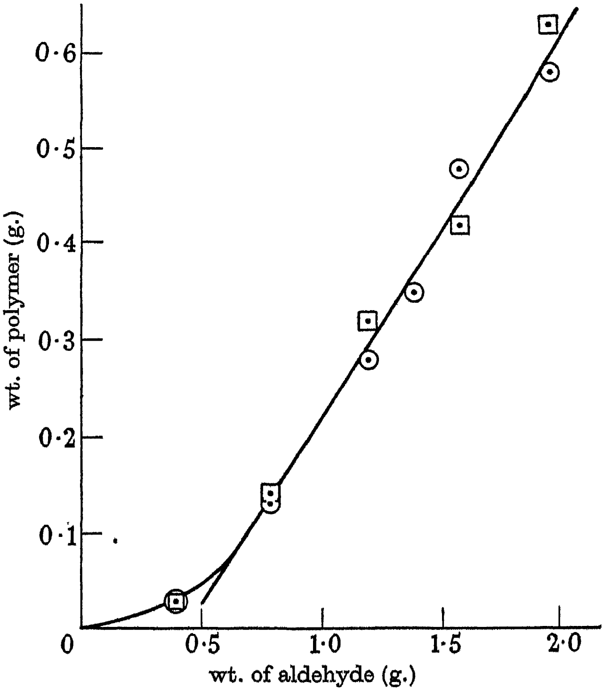


FIGURE 4. A comparison of results obtained by direct condensation of vapour, and by distillation. ⊡ distillation; ⊙ condensation.

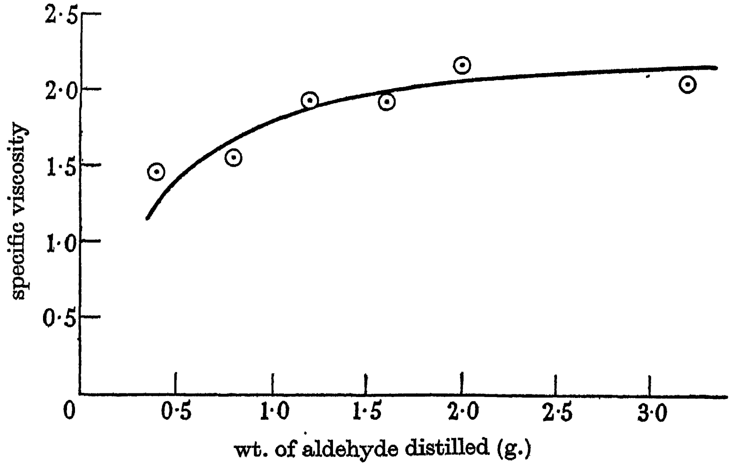


FIGURE 5. The variation in the molecular weight during the course of a distillation. Rate of distillation = 0·080 g./min.

1·6 g. of acetaldehyde were distilled from 0° C to liquid nitrogen at 0·080 g./min. in batches of 0·4 g. with an interval between each; both the yield of polymer and its molecular weight depended upon the length of the intervals (see figure 6). The broken lines on the graphs correspond to results obtained in the same series of experiments, for the distillation of a single batch of 0·4 g. of aldehyde in 5 min. It can be seen that as the interval between the batches of aldehyde was increased, both the yield and the molecular weight of the polymer tended to the values obtained with a single batch of 0·4 g.

TABLE 3

rate of distillation (g./min.)	specific viscosity of polymer solution
0·080	1·93
0·050	2·07
0·114	1·99
0·080	2·12

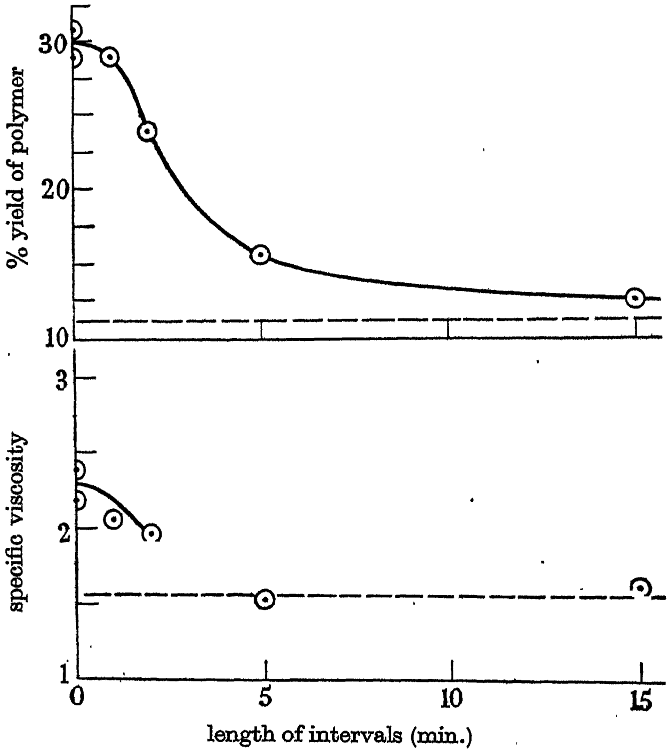


FIGURE 6. The effects of distilling the aldehyde in batches with intervals between them.

The effects of varying the nature of the surface of the receiver

Qualitative experiments were performed to test the effect upon the polymerization, of the temperature of the vessel in which the aldehyde was condensed. When the receiver was at − 119° C which is just higher than the melting-point of the aldehyde, no polymerization was observed; when the receiver was at − 139° C, polymer was formed. These observations confirm that solidification is essential for polymerization.

In certain experiments the inner surface of the receiver was coated with Cetavlon (Norrish & Russell 1947) to see if the nature of the surface had a profound effect upon the reaction. The results quoted in table 4 indicate that substitution of what is effectively a hydrocarbon surface, for one of glass, did not affect the polymerization.

If the receiver contained adsorbed water, the yield of polymer was depressed from about 25 to about 15 %; the presence of adsorbed water may account for the irreproducibility observed in those experiments in which careful attention was not paid to the evacuation and flaming of the receiver. The presence of adsorbed air in the receiver hardly affected the polymerization.

Table 4

surface	% yield of polymer
glass	29·4
Cetavlon	30·0
glass	31·5
Cetavlon	31·9
glass	30·6

The properties of the polymer

The high polymer of acetaldehyde is readily soluble in a number of solvents including diethyl ether, butyl acetate, acetone and acetic acid; it is soluble in carbon tetrachloride and ethyl alcohol. It is insoluble in water and petro-ether, and it swells but appears not to dissolve in benzene. The polymer is rubbery but is without great mechanical strength; materials prepared from monomer containing acetic acid are, however, soft and tacky.

The stability of the polymer

At room temperature the polymer gradually reverts to monomeric acetaldehyde; its stability seems to be intermediate between that of the linear formaldehyde polymers and that of the linear polymers of higher aldehydes, formed at high pressures. The depolymerization in air was followed by weighing a sample of polymer kept in a thermostat. The experiments are not exact, since the recorded weights include any monomeric aldehyde retained in the polymer, and acetic acid might well have been formed by atmospheric oxidation of this monomer. In figure 7, the percentage of the original weight remaining at a particular time is shown as a function of the time, for two separate experiments. The reactions had the characteristics of an autocatalytic process; curve A shows that the rate of decomposition reached a maximum at about 50 % decomposition. As reported by Rigby *et al.* (1948), the stabilities of various samples of polymer were different; for example, curve A refers to an experiment at 25° C, and B to one at 35° C with a different sample of polymer. The tendency for decomposition must increase with rising temperature, yet the rate at 35° C was less than that at 25° C. An explanation of the variations in the stabilities of the samples is advanced below.

It was found that polymers made from samples of aldehyde known to contain acetic acid, and polymers to which a little acid had been added, decomposed more quickly than others. This fact is shown clearly by experiments on depolymerization in solution. A solution of 0·4 g. of polymer in 100 ml. of n-butyl acetate was made; to one portion of 25 ml., 0·1 ml. of acetic acid were added. The specific viscosities of the acidified solution and the pure solution were measured at 20° C, and their variation with time is shown in figure 8. It can be seen that acetic acid catalyzed the process of depolymerization.

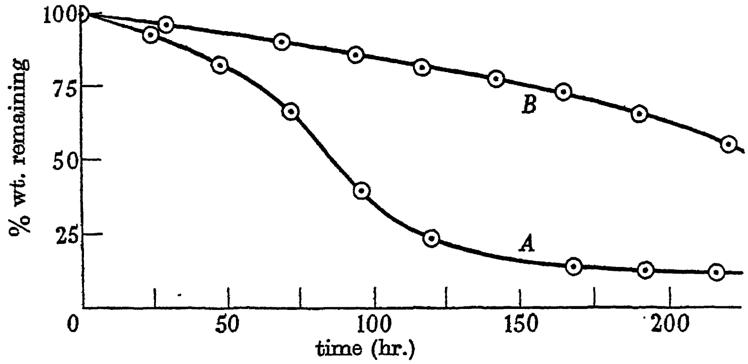


Figure 7. Depolymerization of solid polymer. *A* at 25° C; *B* different sample at 35° C.

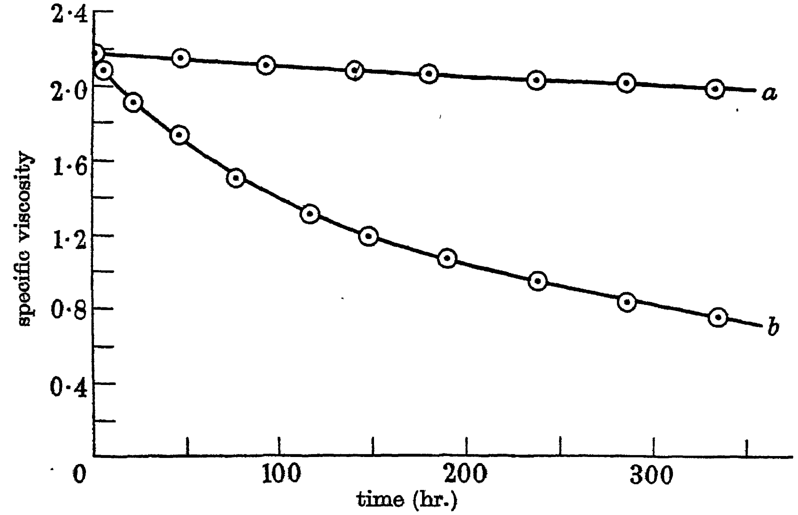


Figure 8. Depolymerization in solution. *a*, polymer alone; *b*, with added acetic acid.

The autocatalysis of the depolymerization in air may well have been caused by acetic acid formed by the oxidation of some of the monomer produced. The differences in the stabilities of various samples of polymer are attributed to variations in their acetic acid content.

Discussion

Letort *et al.* (1947) and Rigby *et al.* (1948) showed that the formation of the high polymer was associated with the freezing of the aldehyde. This result was fully confirmed by the experiments described here, in which the aldehyde vapour was condensed on surfaces at various temperatures. Rigby *et al.* (1948) showed that the yield of polymer was independent of the time for which the aldehyde remained frozen, indicating that reaction occurs during the act of freezing; this conclusion is confirmed by our observation that the addition of alcohol to a sample of frozen aldehyde does not alter the yield of polymer.

The existence of a ceiling temperature above which reaction does not occur, has been demonstrated in the polymerization of formaldehyde (Carruthers & Norrish 1936) and in the co-polymerization of sulphur dioxide and certain olefines, and it is believed that the phenomenon is quite general in polymerizations (Dainton & Ivin 1948). It is necessary to consider the possibility that in the case of acetaldehyde, the melting-point happens to be a ceiling temperature of the type encountered in the formation of polysulphones. This possibility can be eliminated by considering the following results of Rigby *et al.* (1948): (i) supercooled acetaldehyde does not polymerize; (ii) the yield of polymer depends upon the rate at which the liquid is frozen.

It is clear that the actual solidification of the aldehyde, rather than the low temperature, is the governing factor in the polymerization.

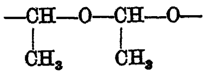
Catalysis of the polymerization

It is considered that no catalyst is required for the reaction. The differences in the behaviour of separate samples of acetaldehyde can be attributed to variations in the concentrations of inhibitors, quite as readily as to variations in the concentration of some unknown catalyst. Among the powerful inhibitors for the reaction are water and alcohol, both of which are possible impurities in the aldehyde. It was thought that peroxides might be catalysts for the polymerization, as had been suggested in the polymerization of certain aldehydes at high pressures (Conant & Peterson 1932). Acetaldehyde certainly forms a peroxide with great readiness (Hatcher, Steacie & Howland 1931) but there is no evidence for it being effective as a catalyst in the polymerization. The effect of phenyl-α-naphthylamine in reducing the tendency for polymerization may be due to the introduction of small amounts of an inhibitor, rather than to the removal of catalytic peroxides.

Acetic acid certainly increases the amount of polymer formed from a sample of acetaldehyde, but it reduces the molecular weight very considerably. The yield of polymer from a sample of acetaldehyde with a small tendency for polymerization, can be raised by the addition of traces of acetic acid, to the same level as for a second, more active, sample of aldehyde with no added acid. The molecular weight of the product from the aldehyde with no added acid, is considerably higher than that of the polymer derived from the acidified monomer. It appears therefore that the different activities of various samples of aldehyde cannot be attributed entirely to differences in their acid content. It will be shown later that the effect of acetic acid can be explained on the assumption that the acid can cause polymer chains to break, there is independent evidence for the occurrence of this process.

The kinetics of the reaction

A summary has already been given of the evidence that the high polymer is a methyl-substituted polyoxymethylene



$$-\underset{\displaystyle CH_3}{\underset{|}{CH}}-O-\underset{\displaystyle CH_3}{\underset{|}{CH}}-O-$$

This structure will be assumed in the discussion which follows.

It is considered that the active centres for polymerization are formed as acetaldehyde is condensed on the cold surface, and that the polymer grows by the addition of monomer molecules from the gas phase. The reaction is similar to the polymerization of gaseous formaldehyde (Carruthers & Norrish 1936) in that it is heterogeneous, involving active centres on the cold surface of the vessel and monomer molecules in the gas phase. The main differences between the two cases are these: (i) for formaldehyde, the surface needs to be at a temperature below about 90° C, but for acetaldehyde it must be below $-123{\cdot}5°$ C; (ii) in the case of acetaldehyde, if a monomer molecule fails to react at its first collision with the cold surface it is likely to be condensed, and yields considerably less than 100 % are therefore to be expected; with formaldehyde, however, it is possible to observe complete polymerization, since the temperature is above that at which the monomer condenses.

The fractional yield of polymer can be expressed as

$$\frac{\text{rate of polymerization}}{\text{rate of supply of aldehyde to receiver}} = \frac{kna}{a} = kn,$$

where k is a constant and n is the number of centres capable of growth. Any factor which alters the magnitude of n must affect the yield of polymer.

The phenomenon of small weights of aldehyde producing smaller percentages of polymer than large weights, is explained by supposing that the active centres are long-lived and that as the aldehyde condenses on the cold surface, their number increases from zero to a stationary value when the rates of formation and destruction are equal. If

$k_1 a$ = rate of production of centres (a being the rate of distillation of aldehyde),
$k_2 na$ = rate of growth of centres,
$k_3 na$ = rate of termination by a reaction between a centre and an aldehyde molecule, of a type different from that leading to growth of the polymer,

the number of centres is governed by the equation

$$\frac{dn}{dt} = k_1 a - k_3 na.$$

Solution of this differential equation and application of the condition that at $t = 0$, $n = 0$ leads to

$$n = \frac{k_1}{k_3}[1 - e^{-k_3 at}].$$

Neglecting consumption of aldehyde in processes other than growth of the polymer, the amount of polymer produced between starting a distillation and stopping it at time T, is given by

$$\int_0^T (\text{rate of growth})\, dt = \int_0^T k_2 na\, dt.$$

Substituting for n, and integrating between the limits shown, leads to

$$w = \frac{k_1 k_2 a}{k_3}\left[T + \frac{e^{-k_3 aT}}{k_3 a} - \frac{1}{k_3 a}\right].$$

This equation indicates that (i) when $T = 0$ the weight of polymer is zero; (ii) the weight increases non-linearly with time until T is large enough for $\frac{e^{-k_3aT}}{k_3a}$ to be negligible, when

$$w = \frac{k_1k_2a}{k_3}T - \frac{k_1k_2}{k_3^2},$$

or

$$w = KaT - K'.$$

If graphs are plotted with (weight) as ordinate and (time) as abscissa, straight lines should be obtained for large values of the time, and they should have slopes proportional to the rate of distillation; the straight lines should not pass through the origin but through the point $(O, -K')$. Figure 3 shows that the straight lines could be drawn to pass through the point $(O, -0{\cdot}14)$ and that the gradients for rates of 0·050, 0·080 and 0·114 g./min. were 0·016, 0·029 and 0·043. The ratios of rate of distillation to gradient of the line are 3·1, 3·6 and 2·7, showing that the gradient is approximately proportional to the rate of distillation.

The low yields of polymer obtained from small weights of monomer might be explained if an essential catalyst less volatile than the aldehyde, tended to be absent from the first fractions, or alternatively if an inhibitor tended to be concentrated in the first fractions. These possibilities are eliminated by the results obtained by condensing aldehyde vapour directly into a vessel cooled in liquid nitrogen (see figure 4); a y catalyst or inhibitor in the monomer would have been condensed at a uniform ate. Exactly the same results were obtained as by the distillation techniqu

The i advanced above, are confirmed by work on the molecular weights of the pol s. In the early stages of the reactions when the number of centres was increas here must have been a preponderance of chains which had been growing for on rt periods; under these conditions the average molecular weight must have w. As the time of distillation was increased, the polymer contained more and olecules which had attained their maximum size, and so the average mole eight should have approached a limiting value. Figure 5 shows that this was ly observed. When the stationary state was reached the average chain leng uld have been defined by the fraction

$$\frac{\text{rate of propagation}}{\text{rate of termination}} = \frac{k_2na}{k_3na}.$$

T ing value of the molecular weight should therefore have been unaffected b ate of distillation. Results presented in table 3 show that this was so ately.

The action of inhibitors

duction in the yield of polymer caused by the presence in the aldehyde of es such as water and alcohol, is explained by the occurrence of a second tion process. It is supposed to involve a centre and a foreign molecule and r at a rate k_4nf, where f is the rate at which the foreign substances enters the er.

The number of centres at the stationary state becomes

$$\frac{k_1 a}{k_3 a + k_4 f}[1 - \mathrm{e}^{-(k_3 a + k_4 f)t}],$$

a quantity which diminishes as f increases. The yield of polymer will be similarly decreased by the presence of the foreign substance.

The limiting value of the average chain length is

$$\frac{k_2 na}{k_3 na + k_4 nf},$$

and this quantity becomes smaller as f is increased. The effect of acetone in reducing the molecular weight of the polymer is therefore explicable.

Spontaneous deactivation of centres

If the centres retained their activity indefinitely on standing in vacuum at the temperature of liquid nitrogen, there should be no differences between the yield and molecular weight of a polymer formed in an experiment in which a given weight of aldehyde was distilled in a single process, and the corresponding quantities for a polymer produced in a distillation which was interrupted periodically. If, however, the centres tended to lose their activity on standing, the percentage yield and the molecular weight of the polymer should have fallen off as the interval between the batches of 0·4 g. was lengthened; these quantities should have tended towards the values obtained in an experiment with a single batch of 0·4 g. Figure 6 shows that these decreases actually occur.

The effect of acetic acid

It has been shown that acidic substances can catalyze the ruptur carbon-
oxygen bonds in formaldehyde polymers (Walker 1944), and experim on the
decomposition of polyacetaldehyde showed that acetic acid can caus e same
effect with this polymer. The effect of acetic acid increasing the yield lyme
(see table 2) may be due to its attacking polymer molecules and splitting t that
each part can act as a nucleus for further polymerization. This would e the
number of centres to increase, and would affect similarly the yield of pol The
average molecular weight of the polymer would be reduced by acetic acid the
acid tends to split the polymer molecules into smaller fragments.

If allowance is made for spontaneous deactivation of centres and also for va-
tion by traces of inhibitors in the monomer as well as for destruction of at
a rate $k_3 na$, the kinetic treatment given previously must be modified. The is
not serious if the additional methods of termination are relatively unir t
compared with the method postulated originally. The decrease in the nu f
centres brought about by these methods of termination might be cancelle
increase caused by traces of acetic acid left in the aldehyde after purificatio

The effect of nitrogen

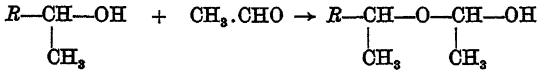
The effects of nitrogen upon the yield and molecular weight of the polym
only partially explained. The decrease in yield might be accounted for if nit

reduces k_2na which is the rate of growth of polymer; this would be the case if the effective concentration (a) of aldehyde near the cold surface is reduced by the layer of nitrogen through which the monomer must diffuse. The striking increase in the molecular weight caused by small amounts of nitrogen is, however, unexplained.

Nitrogen probably has no effect on the kinetic branching introduced by traces of acetic acid left in the aldehyde after purification. Using air instead of nitrogen under conditions which allowed acetic acid to be formed, it was shown that the decrease in molecular weight usually found with acetic acid was actually observed.

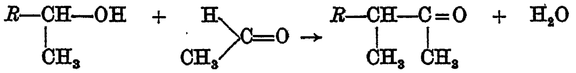
THE MECHANISM OF THE REACTION

It is possible to construct a mechanism for the polymerization, in which the reactivity of a centre is attributed to a terminal hydroxyl group. Such a mechanism would be similar to that proposed by Carruthers & Norrish (1936) for the gas-phase polymerization of formaldehyde with formic acid as catalyst. The propagation step can be written as

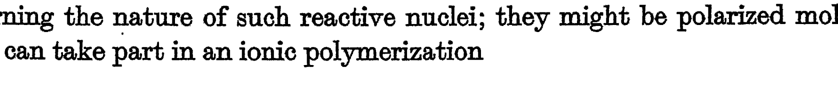


$$R\text{—}\underset{CH_3}{\underset{|}{CH}}\text{—}OH + CH_3.CHO \rightarrow R\text{—}\underset{CH_3}{\underset{|}{CH}}\text{—}O\text{—}\underset{CH_3}{\underset{|}{CH}}\text{—}OH$$

while the alternative reaction of a centre and an aldehyde molecule leading to termination, can be formulated as



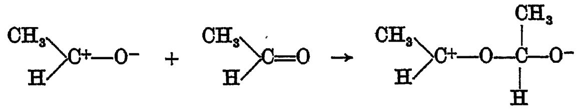
The multiplication of centres caused by acetic acid, can be represented thus



$$\cdots\text{—}\underset{CH_3}{\underset{|}{CH}}\text{—}O\text{—}\underset{CH_3}{\underset{|}{CH}}\text{—}O\cdots + CH_3.COOH \rightarrow \cdots\text{-}\underset{CH_3}{\underset{|}{CH}}\text{—}OH + CH_3.COO.\underset{CH_3}{\underset{|}{CH}}\text{—}O\text{—}$$

It is difficult to explain the powerful inhibiting action of water on this mechanism.

It might be possible that during the solidification of acetaldehyde, two types of nucleus can be formed, the one leading only to crystallization, and the other being capable of acting as a centre of polymerization. Only speculations can be made concerning the nature of such reactive nuclei; they might be polarized molecules which can take part in an ionic polymerization



or alternatively they might be molecules rich in energy which can enter into energy-chain reactions. The latent heat of fusion of acetaldehyde has been given as 0·77 kcal./mole (Cooper 1927) and it is difficult to imagine that the carbonyl bond could be activated by such a small amount of energy.

Evidence for some type of ionic polymerization is provided by the fact that boron trifluoride and acids catalyze the formation of cyclic polymers from liquid acetaldehyde at low temperatures. Reactions induced by these reagents certainly involve polar intermediates and it is quite probable that the formation of the high polymer occurs by a related mechanism. The need for solidification in the production of the high polymer, may be due to the necessity of holding the growing centres fixed in order to prevent premature termination by ring closure. Low yields of polymer were obtained in experiments in which the liquid aldehyde was frozen very rapidly and in experiments in which the monomer was distilled very quickly. These low yields may have been due to the formation of clusters of aldehyde molecules around the active ends of growing centres, preventing them from engaging in further reaction.

Depolymerization is visualized as occurring by the loss of monomer molecules from the ends of polymer chains. Acetic acid catalyzes the decomposition by splitting the polymer molecules into smaller units, so increasing the number of points from which monomer molecules can be shed.

The authors wish to acknowledge a grant from the Research Fund of the Chemical Society. The work was performed while the junior author held a Research Fellowship at Queens' College.

REFERENCES

Bridgman, P. W. & Conant, J. B. 1929 *Proc. Nat. Acad. Sci., Wash.*, **15**, 680.
Carruthers, J. E. & Norrish, R. G. W. 1936 *Trans. Faraday Soc.* **32**, 195.
Conant, J. B. & Peterson, W. R. 1932 *J. Amer. Chem. Soc.* **54**, 628.
Cooper, D. Le B. 1927 *Proc. N.S. Int. Sci.* **17**, 82.
Dainton, F. S. & Ivin, K. J. 1948 *Nature*, **162**, 705.
Hatcher, W. H., Steacie, E. W. R. & Howland, F. 1931 *Canad. J. Res.* **5**, 648.
Letort, M. 1936 *C.R. Acad. Sci., Paris*, **202**, 767.
Letort, M. & Duval, X. 1943*a* *C.R. Acad. Sci., Paris*, **216**, 58.
Letort, M. & Duval, X. 1943*b* *C.R. Acad. Sci., Paris*, **216**, 608.
Letort, M., Duval, X. & Rollin, Y. 1947 *C.R. Acad. Sci., Paris*, **224**, 50.
Norrish, R. G. W. & Russell, K. E. 1947 *Nature*, **160**, 57.
Powell, H. M. 1948 Quoted by Rigby *et al.* (1948).
Rigby, H. A., Danby, C. J. & Hinshelwood, C. N. 1948 *J. Chem. Soc.* p. 234.
Staudinger, H. 1936 *Trans. Faraday Soc.* **32**, 249.
Sutherland, G. B. B. M., Philpotts, A. R. & Twigg, G. H. 1946 *Nature*, **157**, 267.
Travers, M. W. 1936 *Trans. Faraday Soc.* **32**, 246.
Walker, J. F. 1944 *Formaldehyde*, Chap. 7. New York: Reinhold.

Splashes from underwater explosions

By H. Kolsky, J. P. Lewis, M. T. Sampson, A. C. Shearman* and C. I. Snow
Imperial Chemical Industries Ltd.

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 28 *January* 1948—*Revised* 27 *August* 1948—*Read* 3 *March* 1949)

[Plates 8 to 11]

The splashes from the underwater explosions of 1 and 10 lb. charges of P.E. no. 2 and Nobel's Explosive '808' at various depths have been photographed with cine-cameras. The experiments were carried out in a sheltered pond which enabled the surface phenomena to be studied in greater detail than has been done hitherto, and a number of new phenomena have been observed.

Measurements of the initial velocity of rise of the water were carried out with the use of a high-speed cine-camera, and it is shown that for the range of depths

$$0{\cdot}3 < W^{\frac{1}{3}}/D < 4$$

(where W is the weight of the charge in lb. and D is the depth in metres) the initial velocity of rise of the water in m./sec. was given to an accuracy of about 10 % by the relation

$$V = 66(W^{\frac{1}{3}}/D - 0{\cdot}1).$$

From this relation the peak pressure P of the shock wave sent out by the explosion was evaluated in terms of the distance from the charge, i.e. a value was obtained for C in the theoretical relation $P = CW^{\frac{1}{3}}/D$. This was found to be somewhat higher than that obtained by other workers using pressure gauges.

Away from the centre of the dome, the water rose with decreasing initial velocity, and it was found that in calm water the contour of the dome was a cubic of the form

$$H = B\left[\frac{D^2}{D^2+R^2} - C\right],$$

where H is the height of a point at a distance R from the centre of the splash and D the depth of the charge, B and C being constants depending on the nature of the charge.

High-speed photography showed a new phenomenon during the first few milliseconds after detonation, which was provisionally termed the 'crack'. It is seen as a whitening below the surface of the water, which grows at a rate closely corresponding to the arrival at the surface of the shock wave.

The 'black ring' observed around the domes obtained from deep charges was photographed with telephoto lenses and shown to consist of a roughening of the water, resulting from the intensification of ripples already present when the water was at all rough, and from the formation of individual 'spikes', in water which was originally very calm.

Several theories of the formation of domes and black rings are considered, and it seems clear that several mechanisms are operative in producing the phenomena observed. In particular, the enhancement of ripples already present in the water and cavitation beneath the surface appear to occur simultaneously.

The role played by the bubble of gas formed by the explosion is also discussed, and phenomena resulting from its oscillations are described. It is suggested that the central column of water observed with shallow charges is produced as a result of the initial expansion of the bubble, and this theory is shown to account for the observed characteristics of the central column.

INTRODUCTION

In recent years, much has been learnt about the bubble formed by an underwater explosion and the pressure waves generated by it. The behaviour of the water thrown up by such explosions has been studied less extensively, and this paper describes a cinematographic investigation of the phenomena seen above the surface.

---

* Since deceased.

PREVIOUS WORK

Several workers (Lamb 1943; Herring 1947; Taylor 1942; Temperley 1943; Penney & Price 1942; Friedman 1947) have investigated the theory of underwater explosions, and their results may be summarized as follows:

(1) The explosion generates a spherically symmetrical pressure wave in the water. The pressure due to this wave rises to a maximum value in a few microseconds and then decays in a few tens of microseconds to a small fraction of its maximum value; during the first part of this decay the pressure decreases with time exponentially, but later it ceases to do so. For the explosives investigated the pressure wave contained about one-quarter of the total energy of the explosion.

(2) The explosion generates a bubble which undergoes a damped radial oscillation of considerable amplitude. Thus the bubble generated by 1 lb. of T.N.T. detonated at a depth of 2 m. reaches a maximum radius of 1·13 m. and then contracts to a minimum radius of 0·29 m., 0·18 sec. after detonation.

(3) Each time the bubble radius passes through a minimum a pressure wave is sent out, the impulse of the first of which is comparable with that of the first shock wave. At the same time it rises by a distance comparable with its diameter. Its shape may depart considerably from the spherical.

(4) Herring (1947) and Friedman (1947) showed that a free surface near the bubble not only decreases the time of oscillation and radius of the bubble, but also repels it. The pressure of a surface does not affect the emission of the initial shock wave.

Hilliar (1919) and several other workers (Wood & Lakey 1942; R. R. L. 1942–3) have investigated the effects of the shock wave and have described the general appearance of the splash from charges at moderate and at great depths as seen from above the surface.

The phenomena they observed were as follows:

For moderate depths the surface above the charge is seen to rise into a white dome of broken water (such domes are shown in figures 2, 3 and 4, plates 8 and 9). Some time later this dome is broken through by a rush of gas from the explosion, throwing up plumes of spray which generally appear as broad bush-like eruptions, but sometimes take the form of a single thin spout rising to a great height. The photographs showed the domes to be surrounded by a region of black water now known as the black ring. When the charges were fired at great depths, the surface above them did not rise at all, and there was no sign of the water being broken up; on the contrary when the water was smooth the surface appeared darkened 'as still water is darkened by a catspaw of wind'.

Hilliar also mentioned the similarity of the formation of the dome to the Hopkinson effect (Hopkinson 1914), and pointed out that if the water rose from the surface as soon as it was submitted to a tension owing to the shock wave being reflected as a tension wave, its initial velocity would be given by $2P/(\rho c)$, where P is the maximum pressure in the incident shock wave, c is the velocity of the wave and ρ the density of the water. Such reflected waves were observed by Hilliar and others.

EXPERIMENTAL ARRANGEMENTS

The experiments were carried out in a pond at Staines, about 200 m. long, 80 m. wide and 11 m. deep. Approximately spherical charges of Nobel's explosive no. 808 or of P.E. no. 2 ('P.E.') weighing 1 or 10 lb. were wrapped in cloth, suspended from a wire rope across the pond, and initiated with a no. 8 electric detonator and a no. 6 C.E. pellet. The splashes were photographed at speeds from 16 to 2000 frames/sec. with Kodak magazine, special and high-speed 16 mm. cine-cameras; 64 frames/sec. was the usual speed. The best definition was obtained with 'Kodachrome' colour film when the light was intense enough; otherwise panchromatic film was used, with or without a red filter.

It was found necessary when photographing at more than 1000 frames/sec. with the Kodak high-speed camera to provide artificial illumination by means of flares. The flares employed were of the 'Safe Photoflash' type designed by Professor Finch and employed in aerial photography. These were fired some milliseconds before the charge by means of a system of relays, so that they had reached their maximum brilliance when the charge was detonated.

The films were measured either with a travelling microscope or by projection, and when measuring the heights of splashes, correction was made for the position of the camera.

RESULTS

(1) *General appearance of splashes*

The appearance of the splashes at various times after detonation of the charge is shown in figures 1, 2 and 3. These are enlargments of photographs taken at 64 frames/sec., and the times after detonation to which they correspond are shown under each picture. For a very shallow charge (e.g. 10 lb. at 25 cm.) the first phenomenon to be observed is a cylindrical, apparently hollow splash up the centre of which there later rises a tall narrow column of water. When the charge is 50 cm. deep (figure 1), only vestiges of this apparently hollow cylinder can be seen; the first phenomenon to be observed is a dome from which the water rises in a tall narrow spout. As the depth is increased, this spout becomes wider and apparently denser; the splash is then widest about one-half to three-quarters of the way up. At 1 m., the spout appears an appreciable time after detonation, and the maximum height it reaches is less than for shallower charges. As the depth is further increased, the dome reaches almost its maximum height before a spout of water rises through it, whilst at a depth of 4 m. (figure 2) the dome is collapsing before the bubble breaks through it to form plumes.

Comparison of the results for 1 and 10 lb. charges showed that the type of splash for a 1 lb. charge at depth D is similar to that for a 10 lb. charge at depth $D \times (10)^{\frac{1}{3}}$, as might be expected from considerations of dynamical similarity. Figure 3 shows the splash from a 1 lb. charge at 2 m.

When a 1 lb. charge was placed just above the surface, a thin narrow spout was observed, which is presumably formed when a depression in the water made by the explosion closes up. This is like the formation of a spout from the hole caused by a drop falling into water (Worthington 1908).

On the few occasions when the water was exceptionally calm, it was found that the domes obtained from charges at intermediate depths ($W^{\frac{1}{3}}/D$ between 1·0 and 0·3) had quite smooth surfaces, at any rate in the earlier part of their lives; this can be seen from figure 4, which shows the first three frames of the film taken at 64 frames/sec. of the dome from a 10 lb. charge at 3 m. The domes obtained from charges at very much greater depths ($W^{\frac{1}{3}}/D < 0{\cdot}25$) were, however, found to be spiky even when the water was perfectly calm.

Photographs taken with the camera a metre or so above the water surface showed the domes to be surrounded by black rings, the outer edges of which were not very sharp when the water was rough. In calm water the edges were considerably sharper and the rings generally appeared blacker than at other times (see figure 4).

The appearance of these black rings in the cine-records was found to depend on the position of the camera, and on the direction of illumination. Thus in photographs taken from a tower 120 ft. high and 360 ft. from the charge, the black ring was very much less pronounced and did not appear to have any very definite outer edge. Figure 5, plate 9, shows the dome from a 10 lb. charge at 5 m. taken from this position. On one occasion when the water was quite calm and the area around the charge was already darkened by reflexions from the bank, the film showed no black ring, the water surrounding the dome being, if anything, lighter than elsewhere.

(2) *Height of splashes*

The height of the splash is plotted against time for 10 lb. charges of Nobel's Explosive no. 808 in figure 12*a*, and for 1 lb. charges in figure 12 *b*. It may be seen from the graphs for shallow splashes that the splash is subjected to a much greater deceleration than that due to gravity.

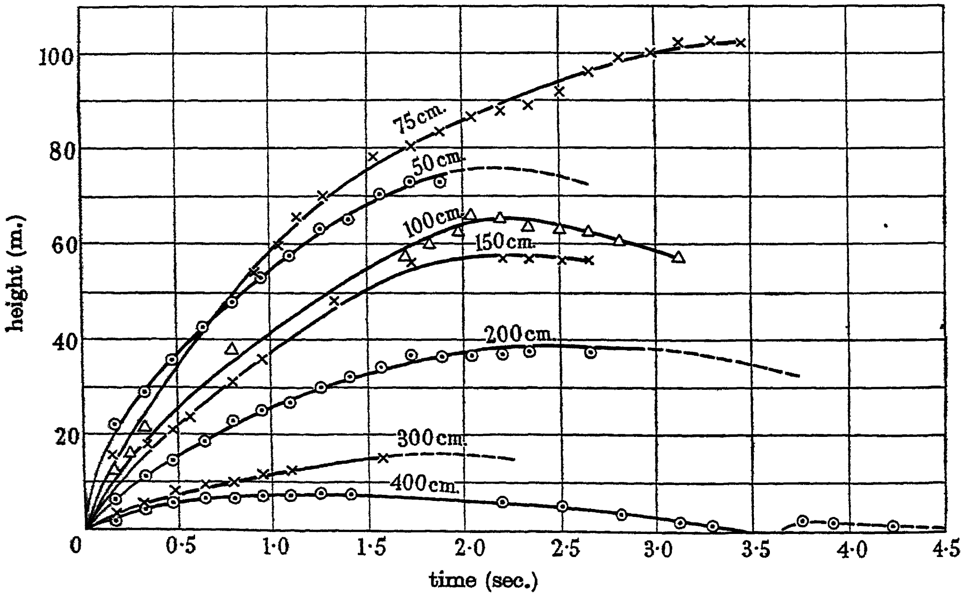


FIGURE 12*a*. 10 lb. unenclosed spheres '808'. Height plotted against time for various depths of detonation.

(3) *Initial velocity of the water*

(*a*) *At the centre of the splash*

It was found that this deceleration was so great for shallow charges that satisfactory values of the initial velocity could not be determined directly from graphs of height against time such as figures 12 *a* and *b*. The ratio height/time was therefore plotted against time, and since we have

$$\frac{s}{t} = V - \tfrac{1}{2}ft,$$

where s is the height and f the deceleration, extrapolation to zero time gives the desired value of the initial velocity V, provided that f does not change too rapidly.

Even when determined in this way the results were rather uncertain, and a number of splashes were therefore photographed with the Kodak high-speed camera at speeds of the order of 500 pictures/sec. This enabled much more satisfactory determinations to be made.

The results obtained in this way are given in table 1 and plotted in figure 13 against $W^{\frac{1}{3}}/D$, where W is the weight of the charge in lb. and D the depth in metres. The

TABLE 1. INITIAL VELOCITIES CALCULATED FROM HIGH-SPEED FILMS

explosive	depth (cm.)	velocity (m./sec.)	explosive	depth (cm.)	velocity (m./sec.)
10 lb. charges			1 lb. charges		
'808'	50	316	'P.E.'	25	236
'P.E.'	100	136·1	'808'	50	135
'808'	200	63·5	'P.E.'	60	94
'P.E.'	300	43·2	'P.E.'	150	30
'808'	700	12·3	'P.E.'	300	14·5

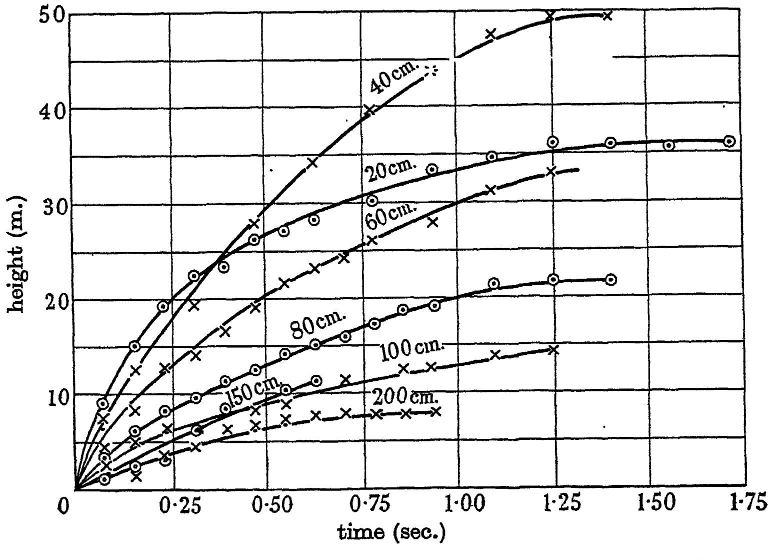


FIGURE 12*b*. 1 lb. unenclosed spheres '808'. Height plotted against time for various depths of detonation.

results given by '808' and 'P.E.' were in good agreement, these explosives having similar characteristics, except that the heat of explosion is somewhat less for '808' than for 'P.E.'.

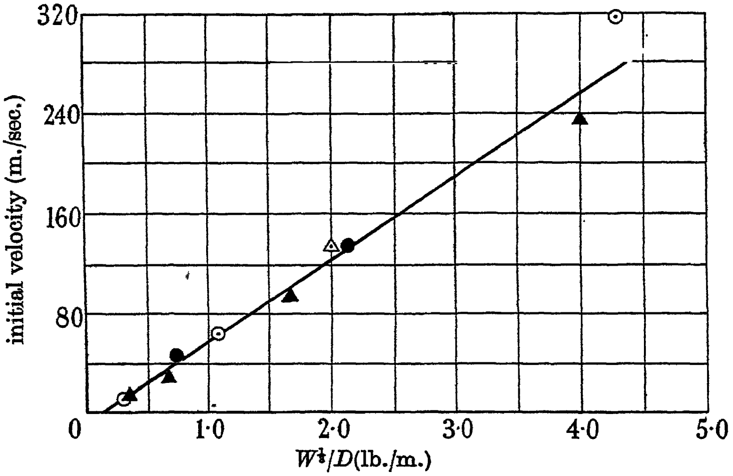


FIGURE 13. Velocity of rise plotted against $W^{\frac{1}{3}}/D$. ⊙ 10 lb. '808'. ● 10 lb. 'P.E.' no. 2. △ 1 lb. '808'. ▲ 1 lb. 'P.E.' no. 2.

It will be seen that over the range $0{\cdot}3 < W^{\frac{1}{3}}/D < 4$, the velocity V is given by the relation

$$V = 66[W^{\frac{1}{3}}/D - 0{\cdot}1], \tag{1}$$

with an accuracy of about 10 %.

For $W^{\frac{1}{3}}/D > 4$, the initial velocities were found to be less than those given by this relation, but owing to the irregular surface of domes at these depths, it was found impossible to measure these velocities with sufficient accuracy to show what other relation might be obeyed.

The mean values of the initial velocities of 1 lb. charges of 'P.E.' at these greater depths were found to be as follows:

depth (m.)	3	4·5	5·0	5·5	6·0	6·5
initial velocity (m./sec.)	14·5	9·0	6·5	5·5	4·5	2·5

The velocity fell to zero at a depth of 7 to 8 m., the white dome disappearing at about the same depth (the 'critical depth'). In rough water, however, small areas of whiteness could often be seen at depths up to 9 m. At greater depths all that could be seen was a circular area of blackness as though the black ring had expanded inwards and obliterated the dome.

(b) *Away from the centre*

The initial velocity of rise away from the centre of the dome could only be measured with any degree of accuracy when the surface of the dome was smooth, i.e. for charges in calm water when $W^{\frac{1}{3}}/D$ had values between 1·0 and 0·3. Under these conditions it was found that the initial velocity at any point except near the edge of the dome obeyed the relation

$$V = \frac{\bar{V}(\cos^2\theta - A)}{1 - A}, \tag{2}$$

where θ is the angle with the vertical of the line joining the point on the surface to the charge, $\overline{V}$ is the velocity of rise at the centre of the dome and A is a constant depending on the weight and depth of the charge.

A short time, t, after detonation, the height H of any part of the dome will be given by Vt. So that, expressing the shape of the dome in terms of D, the depth of the charge, and R, the radial distance from a point on the surface vertically above the charge, we have

$$H = \frac{Vt}{1-A}\left(\frac{D^2}{D^2+R^2}-A\right). \tag{3}$$

Figure 14 shows the comparison between the curve

$$H = 2{\cdot}92\left(\frac{3^2}{3^2+R^2}-3{\cdot}15\right), \tag{4}$$

and the measured heights for a 10 lb. charge detonated at a depth of 3 m., 0·08 sec. after detonation; it may be seen that fairly good agreement has been obtained.

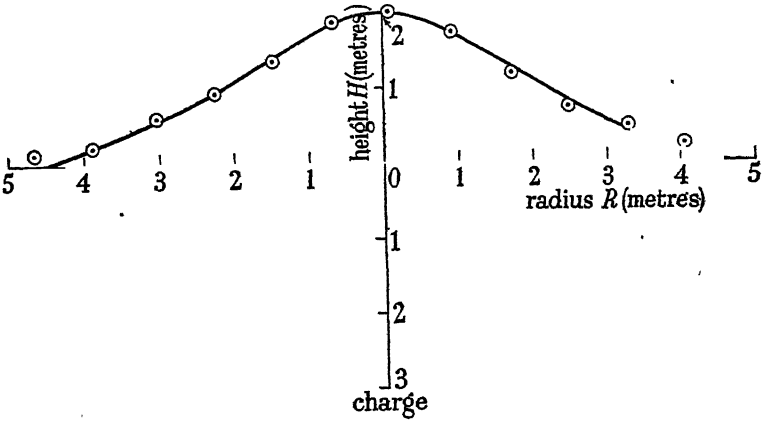


FIGURE 14. Shape of dome in calm water. 10 lb. at 300 cm. (after 0·078 sec.). —— curve of 26·3 [1/(9 + R^2) − 0·35] plotted against R. ⊙ measured from film of 10 lb. at 300 cm.

(4) *Deceleration of water and size of drops*

The initial deceleration of the water in the splash may be calculated approximately from the s/t against t curves mentioned in § 3. The results so obtained are given in table 2.

TABLE 2. INITIAL RETARDATION

depth (cm.)	retardation (m./sec./sec.)	depth (cm.)	retardation (m./sec./sec.)
10 lb. charges		1 lb. charges	
50	2500	30	600
75	600	40	650
100	350	60	400
150	300	70	300
200	300	80	300
300	250	90	250
400	180	100	200
		175	100
		200	100

The high initial velocity and deceleration of the drops present difficulties in the measurement of their size, since the Reynolds numbers met with are much higher than any for which the deceleration has been calculated or indeed observed previously. Measurements were therefore made where possible of the terminal velocity of fall of clouds of drops in the splashes, and in a number of cases this was found to be 8 m./sec. In no cases was it higher than this, but it was sometimes less. Flower (1928) has shown that the maximum velocity of fall of a drop is 8·1 m./sec. and that this is reached by drops of 5 mm. in diameter, larger ones becoming deformed or broken before they can reach a higher velocity. It is, therefore, unlikely that any drops much larger than 5 mm. in diameter persist long after the splash begins to rise.

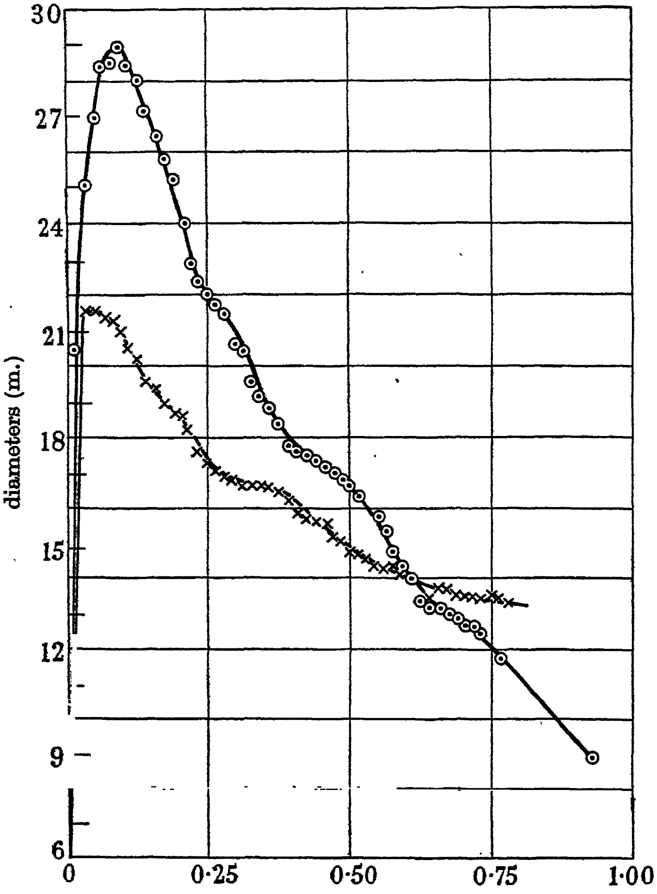


Figure 15. Diameters of domes plotted against time.
× 10 lb. at 300 cm. ⊙ 10 lb. at 1000 cm.

(5) *Diameter of dome and black ring*

The films taken at 64 frames/sec. showed that the domes did not reach their maximum diameters until the second or third frame, and then started to contract slowly. The black ring on the other hand did not normally appear at all until the second frame, and did not subsequently change very much in diameter (see figure 3, plate 8).

Figures 15 and 17 show the growth and decay of the dome diameters for 10 and 1 lb. charges, and in figure 17 the behaviour of the black ring is also shown. These curves are characteristic of all charges in rough water for the range $0{\cdot}2 < W^{\frac{1}{3}}/D < 2$; the time scale increasing proportionately to $W^{\frac{1}{3}}$ for charges of different mass at corresponding depths.

In figure 15 the growths of domes from 10 lb. charges at 3 and 10 m. are compared; it may be seen that the dome contracts very much more rapidly for the deeper charge and this was found generally to be so.

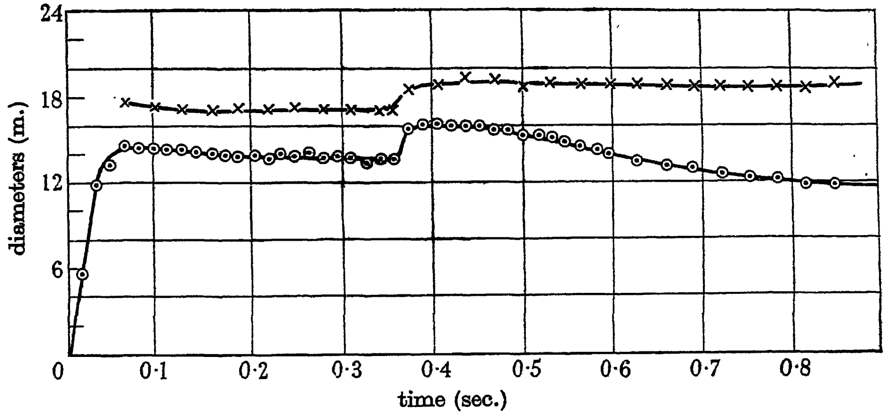


FIGURE 16. 10 lb. at 300 cm. in very calm water. Diameters of dome and black ring plotted against time. ⊙ dome. × black ring.

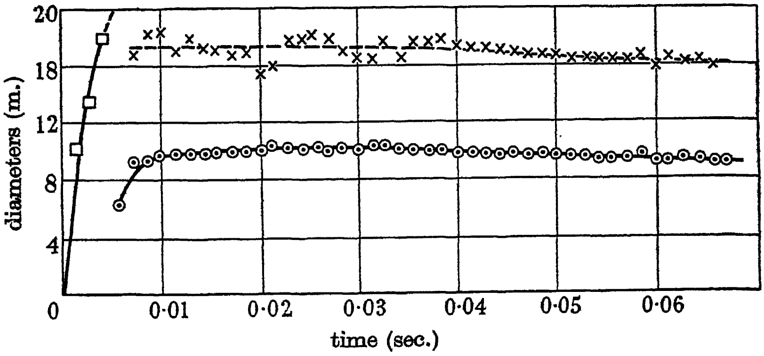


FIGURE 17. 1 lb. at 500 cm. taken with high-speed camera. Diameters of dome and black ring plotted against time. ⊡ crack. ⊙ dome. × black ring.

If it is assumed that the contraction of the dome corresponds to the water falling back to the level of the undisturbed surface, the rate of contraction of the dome diameter should give an indirect method of determining the initial velocities at the outer edge of the dome. Since the velocity of rise is very small, the drops will be decelerated mainly by gravity, so that the time for a drop to rise and fall back into the water, t, is related to the initial velocity, u, by the relation

$$u = \tfrac{1}{2}gt. \tag{5}$$

The shape of the decay curve should therefore be similar to the profile of a vertical section of the dome near its edge if the time axis is multiplied by $\frac{1}{2}g$ and taken to correspond to initial velocity of rise. An examination of the early part of the decay curves indicates that this is at least approximately true, the more gradual slope at the edges of the domes from deeper charges corresponding to the sharper drop in these curves.

In calm water the diameters of the domes were found to be considerably smaller than when the water was at all rough, and figure 16 shows the curve for the dome and black ring of a 10 lb. charge at 3 m. under these conditions. On comparing this with figure 15, it can be seen that the maximum diameter of the dome is considerably less, and that one-third of a second after detonation there is a sharp rise in the diameter of the dome and the black ring. This rise in dome diameter brings it to the value found for the same time in rough water shown in figure 15 and the subsequent decay of the dome diameter may be seen to be the same in the two cases.

This sudden rise in the diameters of the domes in calm water, as well as the quite definite humps seen in figure 15 for charges in rough water, may be correlated with the secondary shock waves sent out by the oscillations of the bubble and are dealt with in detail in a later section.

As the depth of the charges was increased, the maximum diameters of the dome and black ring were found to increase up to depths corresponding to a 1 lb. charge at 5 m. At greater depths the diameters began to contract, the dome diameter decreasing to zero at the 'critical depth' which was about 8·5 m. for a 1 lb. charge.

In tables 3 and 4 these maximum diameters are given for 1 and 10 lb. charges at various depths. The values of $P \cos \theta$ are also given, where θ is the angle with the vertical of the line joining the charge to the edge of the dome and black ring and P is the corresponding maximum pressure at the edge. The calculation of P is discussed in a later section.

It will be seen that for $W^{\frac{1}{3}}/D > 1$ these products are approximately constant, but show a tendency to increase for the deeper charges. The ratio between these products for the edge of the dome and the black ring is given in the last column of the tables and may be seen to be approximately 2.

(6) *Fine structure of the black ring*

In order to elucidate the nature of the black ring a number of photographs were taken at 64 frames/sec. with telephoto lenses. These gave the following results which are illustrated in figures 6 and 7, plates 9 and 10.

A. *In very calm water* the first and second frames show a number of small hillocks (figure 6), from which continuous jets of water then rise to a height of about 10 cm. and then break into drops a few millimetres in diameter.

Some indication of the reason for the formation of these individual spikes was given by an experiment in which pieces of string were hanging down into the water. The cine-film of the dome showed small jets of water climbing up the string even in the area of the black ring, indicating that the formation of these spikes is facilitated by inhomogeneities in the surface of the water.

B. *For rough water* the appearance of the black ring is shown in figure 7. It may be seen that the waves already present in the water are markedly increased in amplitude; the crests of those nearest the centre of the dome being broken up into drops. The presence of individual spikes can also be clearly seen, but these too, tend to be concentrated at the crests of waves.

TABLE 3. MAXIMUM DIAMETERS OF DOMES AND BLACK RINGS FOR 1 LB. CHARGES

depth (cm.)	max. diameter (m.)		$P\cos\theta$ dome (lb./sq.in.)	$P\cos\theta$ black ring (lb./sq.in.)	ratio of pressures
	dome	black ring			
70	6·5	9·1	445	232	1·92
80	6·6	10·0	489	233	2·10
90	7·0	10·2	474	237	2·00
100	7·3	10·2	492	260	1·89
125	7·1	12·6	618	215	2·88
150	8·3	12·6	545	252	2·16
175	8·8	13·6	548	250	2·19
200	9·2	13·8	555	273	2·03
250	9·4	14·2	618	310	1·99
300	10·4	18·4	585	226	2·59
400	9·4	17·2	740	312	2·37
450	10·1	17·6	690	325	2·12
500	10·2	18·6	688	316	2·18
550	12·2	17·2	575	368	1·60
600	10·4	—	668	—	—
650	9·2	—	740	—	—
700	9·9	20·4	668	321	2·08

TABLE 4. MAXIMUM DIAMETERS OF DOMES AND BLACK RINGS FOR 10 LB. CHARGES

depth (cm.)	max. diameter (m.)		$P\cos\theta$ dome (lb./sq.in.)	$P\cos\theta$ black ring (lb./sq.in.)	ratio of pressures
	dome	black ring			
50	11·0	15·2	248	130	1·91
75	11·7	17·0	326	155	2·10
100	14·0	19·6	302	155	1·95
150	14·9	19·4	395	235	1·68
200	15·2	—	490	—	—
250	20·8	28·4	334	171	1·95
300	22·2	30·0	342	193	1·77
350	22·6	—	380	—	—
400	24·0	—	378	—	—
500	23·8	—	454	—	—
700	26·6	—	463	—	—
1000	29·0	—	487	—	—

From the foregoing experiments it is concluded that the 'blackness' of the black ring arises from the formation of 'spikes' and the roughening of the surface of the water. This reduces the amount of light from the sky reflected to positions where the line of view makes a small angle with the horizontal.

The water in the dome, on the other hand, is so completely broken up that the incident light is scattered equally in all directions, giving it a white appearance.

(7) *The crack*

The first frame of the photographs at 64 frames/sec. sometimes showed a blurred circular area of whiteness considerably greater in diameter than that observed in the second and subsequent frames. In order to investigate this phenomenon, and to study in greater detail the early growth of the dome and the appearance of the black ring, some experiments were carried out, using the Kodak high-speed camera.

As has already been mentioned, it was found necessary to use flares for this work and these were suspended vertically above the charge, the camera being 60 ft. up a tower, 360 ft. away from the charge.

The films from the high-speed camera showed that the first phenomenon to be observed above the surface, after the charge had been fired, was a bright circular patch which expanded rapidly over the surface of the water, becoming darker in colour so that it could no longer be seen a few msec. after detonation. The effect is shown in figure 8, plate 10, which gives the first three frames from a 1 lb. charge at a depth of 2 m. taken at 2000 frames/sec. (the vertical streaks on the photographs are due to small particles from the flare impinging on the water).

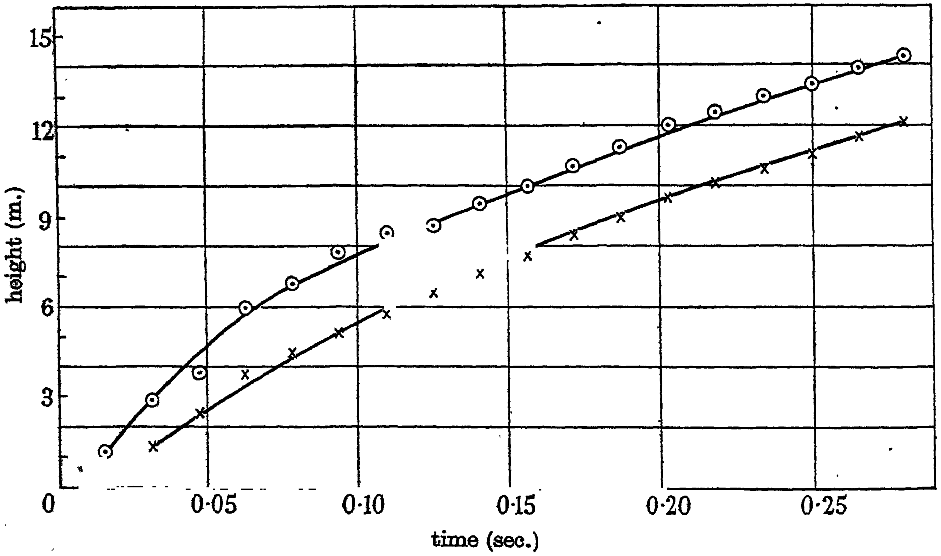
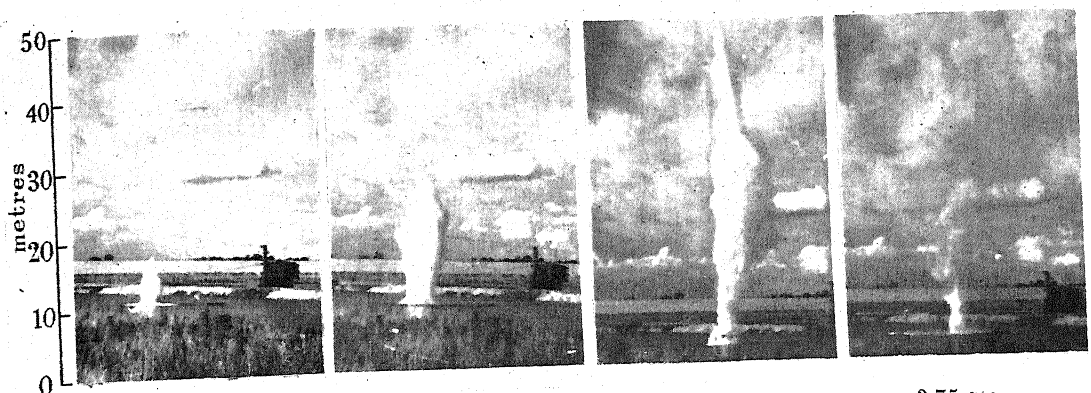


FIGURE 18. 1 lb. unenclosed sphere 'P.E.' no. 2 at 50 cm. Dye vertically above charge. ⊙ top of splash. × top of dye.

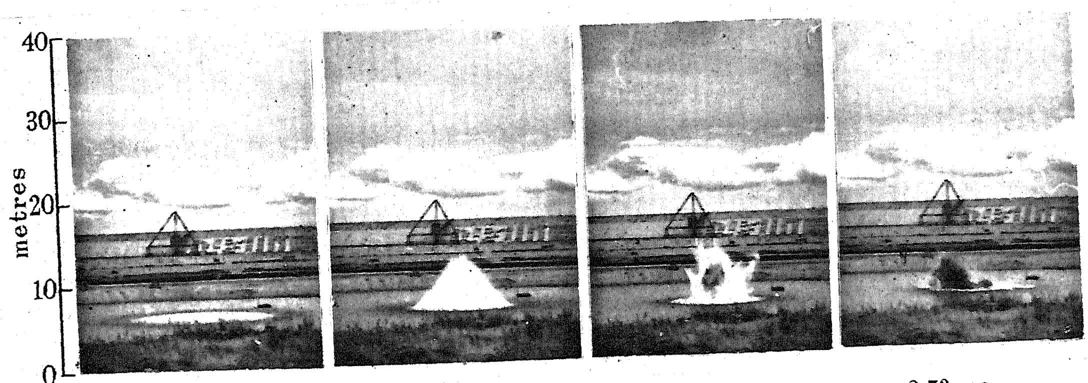
The edge of the white area is quite sharp, and the rather blurred translucent nature of the whiteness indicates that the phenomenon is taking place beneath the surface of the water. This was confirmed by an experiment carried out in very calm water, when the reflexion from the bank remained quite sharp in the region of the phenomenon, showing that the surface of the water had remained undisturbed. In view of this fact, and by analogy with the formation of scabs from solids by explosives, the phenomenon has provisionally been called the 'crack'.

Figure 17 shows the growth in diameter of the 'crack', the dome and the black ring, for a 1 lb. charge at a depth of 5 m. taken at 706 frames/sec.; the outer edge of



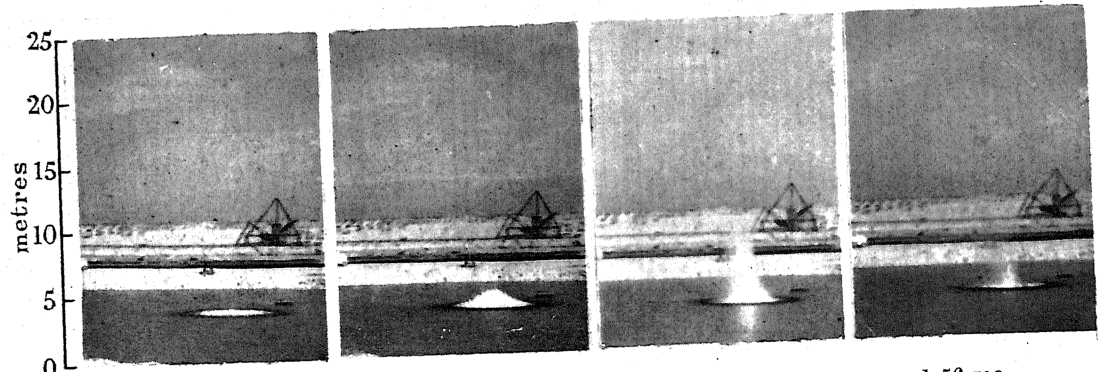
0·03 sec. 0·24 sec. 1·17 sec. 3·75 sec.

FIGURE 1



0·03 sec. 0·77 sec. 1·14 sec. 2·73 sec.

FIGURE 2



0·02 sec. 0·11 sec. 0·92 sec. 1·56 sec.

FIGURE 3

For descriptions of plates see p. 401

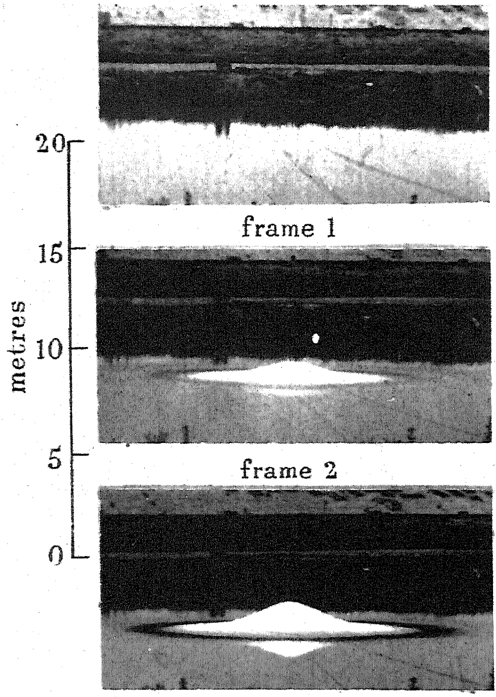


FIGURE 4

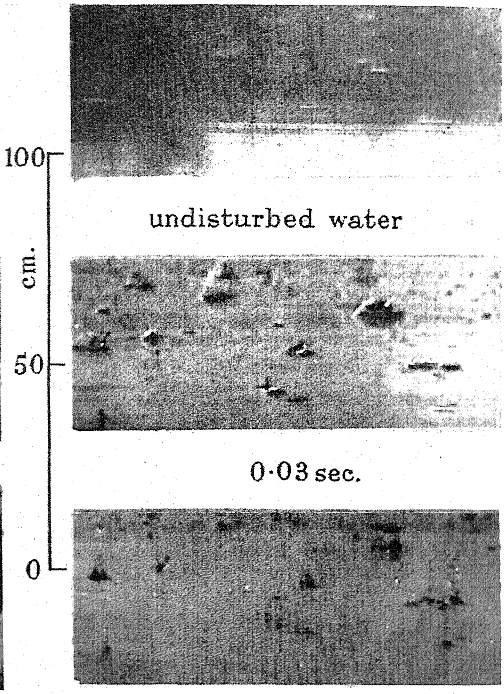


FIGURE 6

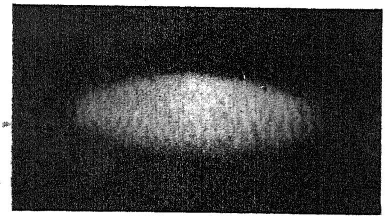
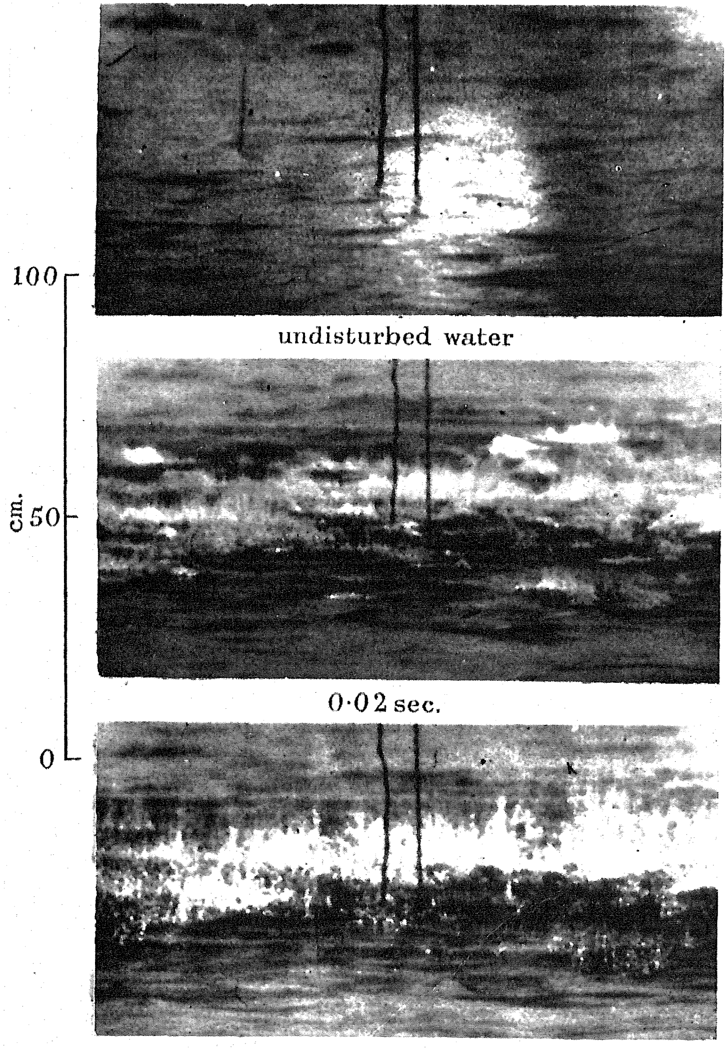


FIGURE 5

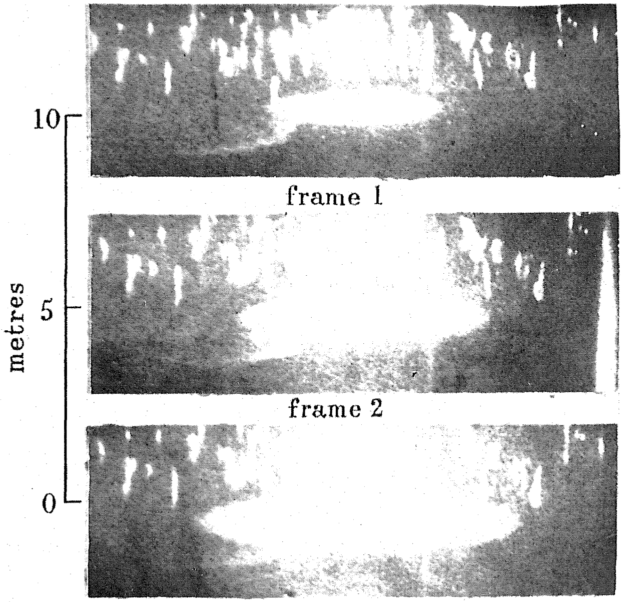


undisturbed water

0·02 sec.

0·05 sec.

FIGURE 7

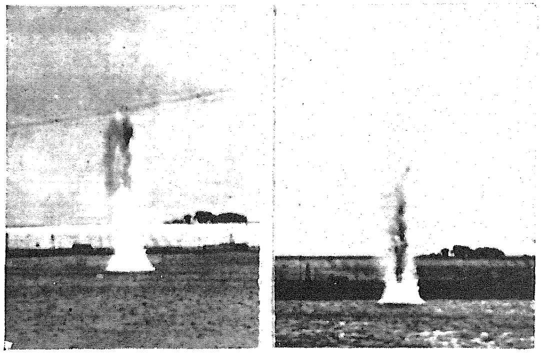


frame 1

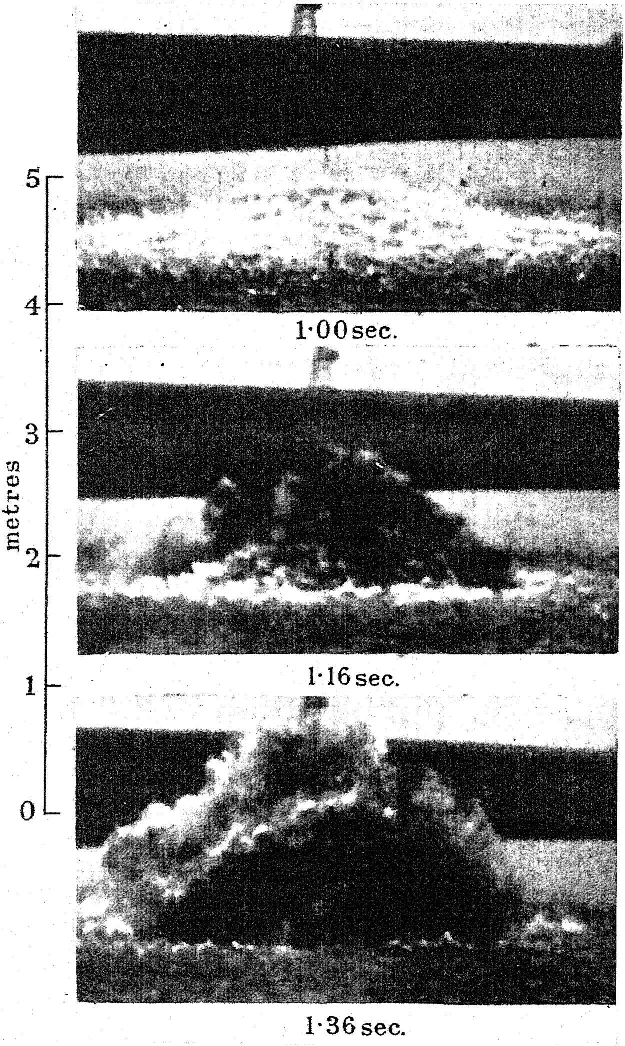
frame 2

frame 3

FIGURE 8



a FIGURE 9 *b*



1·00 sec.

1·16 sec.

1·36 sec.

FIGURE 10

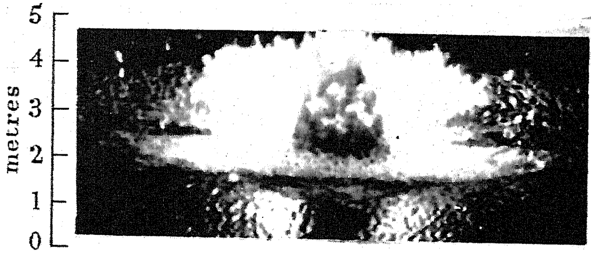


FIGURE 11

the black ring was very ill-defined for the first few frames and the measurements of its diameter consequently show a large scatter.

The radial distance from the charge to the edge of the 'crack' was found to increase linearly with time at a velocity of 1540 m./sec. This is extremely close to the velocity of sound in water and indicates that the appearance of the 'crack' corresponds to the arrival of the shock wave at the water surface.

(8) *Effects due to the bubble*

Most of the effects so far described are associated with the pressure wave sent out by the explosion and reflected at the free surface of the water. In addition to these, however, other phenomena occur which appear to be due to the large bubble of gas generated by the explosion, and which in the case of the shallowest charges completely mask the dome produced by the shock wave.

The phenomena may be conveniently considered under three headings:

(*a*) The central spout which is observed with shallow charges. This is the effect of the initial *expansion* of the bubble.

(*b*) The discontinuities in the rate of decay of dome diameters, and the appearance of subsidiary domes. These phenomena are associated with the subsequent *oscillation* of the bubble.

(*c*) The emergence of the gases when the bubble has *risen* to the surface under gravity.

(*a*) *The central spout*

Though the internal structure of the splashes cannot easily be observed owing to the screening of the cloud of drops around it, it was noticed that in some photographs there appeared to be a central column of water of considerably smaller diameter and higher density than the rest of the splash. Some experiments were therefore carried out in which dye was inserted at various points about the charge and the splash was photographed with 'Kodachrome' colour film.

A magenta dye, known as 'Acid Magenta AS', was used. 1 oz. of the powdered dye was placed in a test tube to which a no. 8 detonator was attached by means of adhesive tape. Preliminary experiments showed that this broke up the tube satisfactorily. This assembly was placed at various points with respect to 1 lb. charges at 50 cm. depth, the two detonators being wired in parallel.

When the dye was placed below the charge no coloration was seen. When it was placed in the surface of the water above the charge, the top part only of the splash was coloured (figure 9*a*, plate 10). When it was placed immediately above the charge, or half way between the charge and the surface, the splash remained white except for a central column which was dyed a deep magenta (figure 9*b*).

The height of the dyed central column was measured at various times and the results are shown in figure 18. From this it will be seen that the central column rose with a velocity comparable with that of the top of the splash. It is believed that whilst the top part of the splash (that shown dark in figure 9*a*) is due to the effect of the shock wave, the central column is due to the expansion of the bubble after the shock wave has left it.

(b) *Phenomena associated with the oscillation of the bubble*

For very deep charges ($W^{\frac{1}{3}}/D < 0{\cdot}2$) a secondary dome was observed to rise from the surface after the original dome has begun to subside. Figure 19 shows the height-time curve for a 1 lb. charge at a depth of 6·5 m. and it may be seen that at both 0·16 and 0·36 sec. after detonation, the water rises after having begun to subside, the effect at 0·36 sec. being the much more marked. Similar effects were observed with shallower charges up to a depth of 4·5 m. although the discontinuity at 0·16 sec. became less marked as a result of the water from the first dome not having reached its maximum height by that time.

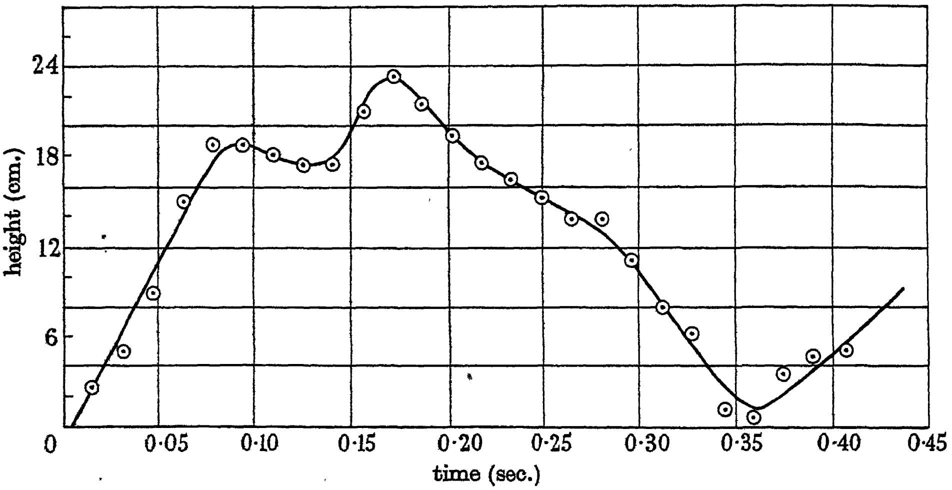


FIGURE 19. 1 lb. at 650 cm.

The theory of the motion of the bubble has been worked out by Lamb, Taylor and others and shows that if the charge is at a sufficient depth for it not to break surface in its original expansion it will reach a maximum diameter and then contract again until it reaches approximately its original diameter. It will continue to oscillate in this way, while at the same time rising under gravity. It may further be shown that during its expanding stages it rises very little, but that the rate of rise increases while it is contracting, and that in the region of its minima it rises very rapidly indeed. It seems reasonable to correlate the formation of the subsidiary domes with the shock waves sent out by the oscillating bubble as it passes through its minima.

Although the pressure pulses sent out by the oscillations of the bubble will decrease in amplitude at each oscillation, as a result of both the energy lost in the preceding pulses and viscosity effects, the bubble will at the same time have risen towards the surface, so that the final effect of the secondary pulses will be similar to that of a smaller charge nearer the surface.

Nobel's Explosive no. 808 may be set off at a lower velocity of detonation (2300 m./sec.) if it is primed with a no. 6 detonator, instead of with a no. 8 detonator and tetryl pellet. Under these conditions the same total chemical energy is released as for ordinary initiation, but the maximum amplitude of the outgoing shock wave is found to be smaller. In this case the initial velocity of rise of the dome is about one-third of that for the explosive detonated normally and the secondary dome effects can be seen much more clearly.

Figure 20 shows the height-time relation for a 5 lb. charge of '808' detonated in this way at a depth of 4 m. It can be seen that the secondary dome appears to begin rising about 0·25 sec. after detonation and reaches a height about equal to that of the original dome. It is of interest that the water in the domes under these conditions is decelerated far less rapidly than for comparable charges detonated normally, indicating that the 'degraded' shock wave has not broken up the water so completely.

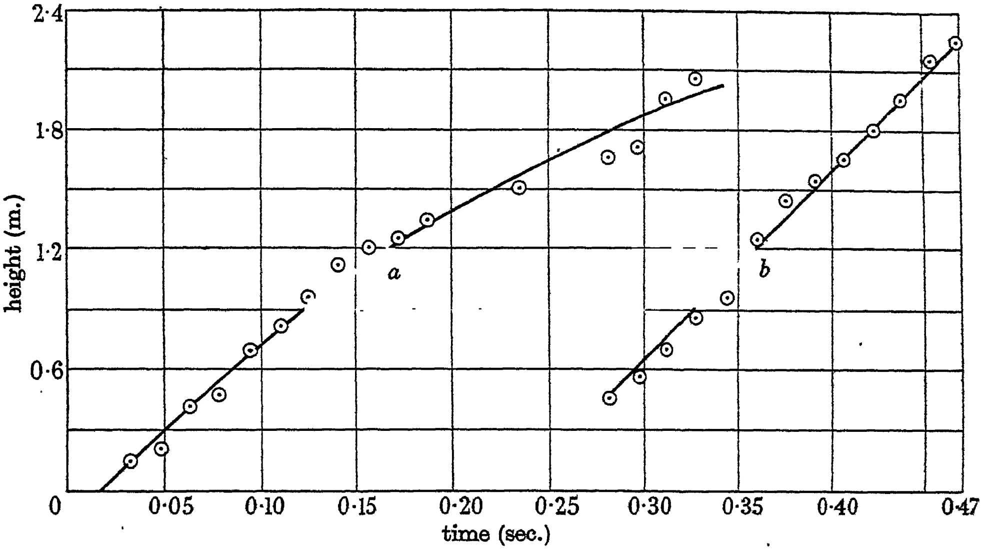


FIGURE 20. 5 lb. '808' at 400 cm. low V.O.D. *a*, original dome. *b*, second dome.

Figure 16 shows the decay of dome diameter for a 10 lb. charge of 'P.E.' no. 2 at 3 m., the experiment having been carried out in very calm water. In this graph a very sharp discontinuity may be seen at 0·375 sec. after detonation, where there is a sharp rise in the diameters of the dome and black ring. When the water was at all rough the effect was never so pronounced, but for all 10 lb. charges a discontinuity in the graphs of dome diameter against time was found to occur between 0·35 and 0·4 sec. after detonation, whilst for 1 lb. charges the discontinuity was at about 0·2 sec.

Here again the effects may be explained in terms of a secondary shock-wave sent out by the bubble during its oscillation, the water that had subsided on the outer edges of the dome being again disturbed by the pulse from the bubble when it passes through a minimum.

From the foregoing results on dome diameters and the appearance of subsidiary domes, it would seem clear that the period of oscillation for the bubble from a 1 lb. charge is about 0·2 sec. and for a 10 lb. charge about 0·4 sec., the deeper 1 lb. charges often showing more than one complete oscillation.

(*c*) *The emergence of the bubble*

The final phenomenon which is observed from deep charges is an upheaval of the water due to the emergence of the gases in the bubble. With charges where $W^{\frac{1}{3}}/D < 0{\cdot}3$ this may be clearly seen after the other effects have begun to subside, and figure 10, plate 11, shows three stages in the emergence of the bubble for a 1 lb. charge at 5 m. As the depth is increased the time after detonation of the appearance of the gases increases and in figure 21 the times of appearance are plotted against depth for charges between 3 and 8 m.

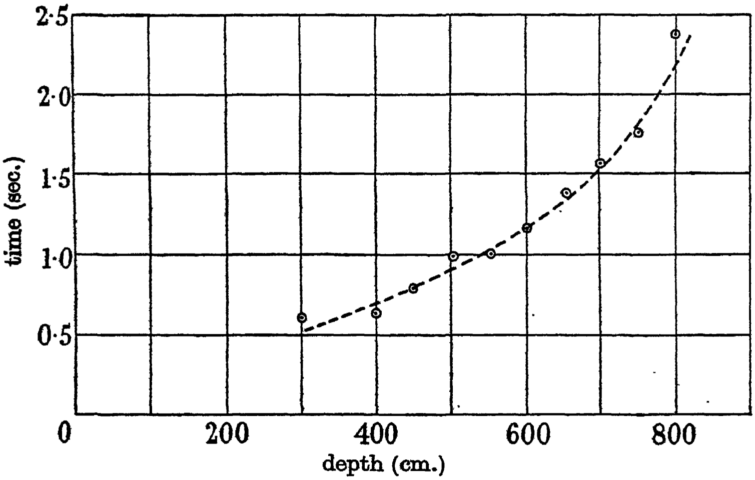


FIGURE 21. Time of appearance of bubble for deep 1 lb. charges.

It is difficult to interpret this curve theoretically, for as Taylor (1942) has shown, the rate of rise of the gas-bubble under gravity varies considerably as the bubble oscillates, so that with deep charges, where several oscillations have taken place, the total time of rise will not be a simple function of the depth. It will be further complicated by the effects of the surface already mentioned. We might, however, expect a periodic step-like variation from a smooth curve, and the graph gives some indication of this.

Taylor has also shown that the bubble is unstable during its contracting stages, and in a few of the experiments with deep charges there was evidence of two or more separate disturbances, showing that the bubble had broken up.

For shallow charges also, a secondary upheaval was found to take place some time after the appearance of the splash. This produced plumes around the central column of water and the time after detonation at which this occurs for charges at various depths is given in table 5.

TABLE 5. TIME OF APPEARANCE OF 'BUBBLE'

			10 lb. charges					
depth (cm.)	50	75	100	150	200	300	400	
time (sec.)	0·78	0·59	0·61	0·58	0·61	0·55	0·88	
			1 lb. charges					
depth (cm.)	20	30	40	50	60	70	80	90
time (sec.)	0·53	0·47	0·45	0·37	0·37	0·36	0·36	0·33

It will be seen that here the time first decreases with increasing depth. This may be due to the bubble breaking out after being repelled from the surface as suggested by Herring (1947).

DISCUSSION OF RESULTS

(1) *Calculation of pressure from initial velocity*

It may be shown from theoretical considerations that the maximum pressure of the shock wave generated by an underwater charge of weight W at a distance D from the charge is given by

$$P = \frac{CW^{\frac{1}{3}}}{D}, \tag{6}$$

where C is a constant depending on the type of explosive employed.

This relation will hold if the compressibility of water remains constant up to the pressure P.

The Road Research Laboratory (R.R.L. 1942–1943) has discussed the relationship between the initial velocity produced at the water surface by such a shock wave, and the peak pressure of the wave. In the case where the tensile strength of water is comparable with this peak pressure they have derived relations for waves in which the pressure declines linearly and exponentially with time.

When the peak pressure is large compared with the tensile strength both these conditions lead to relations which approximate to

$$V = \frac{2}{\rho c}\left(P - \frac{T}{2}\right), \tag{7}$$

where c is the wave velocity, ρ the density and T the tensile strength of water.

The results for initial velocities shown in figure 13 correspond to pressures which are large compared with the tensile strength of water and hence the expression for the velocity given in equation (1), i.e. $V = 66\,(W^{\frac{1}{3}}/D - 0{\cdot}1)$ must be equivalent to an expression of the form $\frac{2}{\rho c}\left(P - \frac{T}{2}\right)$. Thus in the light of equation (6) we may write

$$\frac{2P}{\rho c} = \frac{66\,(W)^{\frac{1}{3}}}{D}, \tag{8}$$

within the experimental error. By taking $c = 1{\cdot}5 \times 10^3$ m./sec. and $\rho = 1$ g./cm.^{3} in equation (8) we find that C in equation (6) is equal to 23,600 in the units used by the R.R.L. (lb./sq.in. for pressure, lb. for W, ft. for D).

No direct measurements of C in equation (6) have been recorded for '808', and the only value available for 'P.E.' is that of 16,000 in their units obtained by the R.R.L. for charges of the order of 1 oz. in weight (Mumford, private communication). A number of measurements have been made on various scales with T.N.T., and Taylor (1942) quotes the value $C = 4{\cdot}9 \times 10^9$ in c.g.s. units $=$ 15,500 in the R.R.L. units. Both these values are considerably smaller than the result given here.

This discrepancy may be due to a difference in the heat of the explosion in the case of T.N.T. and the considerable difference in the weight of charge used in the case of the R.R.L. It is considerably greater than the deviation of our individual results obtained with the high-speed camera, and the systematic errors seem more likely to give too low than too high a result. Further, this method of measuring the maximum pressure is not affected by the extremely rapid variation in pressure which takes place near the maximum, and which may lead to difficulties with inertia when any form of pressure gauge is used.

It is, therefore, concluded that for charges of 'P.E.' or '808' of weight 1 to 10 lb. the value of C in equation (6) is $23{,}000 \pm 2000$ in the R.R.L. units.

(2) *Theories of the mechanism of dome formation*

The mechanism, by which the shock wave from the detonation of the explosive causes the complex phenomena described in previous sections, is not at all well understood. Three distinct theories have been put forward to account for some of these phenomena, but no single theory appears to explain all of the observed effects. A brief account of each of these three theories is given below.

(*a*) *The layer theory*

The physical basis of this theory, which was first put forward by Hilliar (1919) is the assumption that when water is subjected to very rapid stresses by the passage of a shock wave, it has a fixed tensile strength of several hundred pounds per square inch. Thus, when the head of the outgoing pressure pulse is reflected at the free surface of the water, a reversal in phase occurs and the tension set up results in a 'fracture' below the surface, a layer of water being set up with some of the momentum of the pulse trapped in it (cf. The 'Hopkinson Bar' (Hopkinson 1914)). The next part of the pulse is then reflected at the new surface and a whole series of such layers are sent up until the amplitude in the tail of the pulse is insufficient to cause further rupture.

This theory accounts for the phenomenon previously described as the 'crack', for the relation between depth and initial velocity, and for the existence of a 'critical depth'. It gives no explanation, however, of the phenomenon of the black ring, or of the formation of individual 'spikes'.

One of the present authors (C.I.S.) has suggested an extension of the theory to account for the black ring phenomena. This is that in the region of the black ring, the shock wave is only of sufficient amplitude for one layer to be formed; this layer is rapidly decelerated as a result of the vacuum formed beneath it, so that it falls back into the water, and as a result of the impact, a roughening of the water takes place which is seen as the black ring. The comparatively sharp outer edge of the black ring in calm water is thus taken to correspond to the maximum diameter at which the shock wave is of sufficient amplitude to cause cavitation in the water.

In the region where two or more layers are formed, the second and subsequent layers are not decelerated in this way as there is a vacuum on either side of them; consequently the first and second layers will collide and break up into the drops which give the dome its white appearance.

The fact that the 'crack' is seen to extend considerably beyond the maximum diameter of the subsequent dome, and that the maximum observable diameter of the 'crack' is generally approximately the same as that of the black ring, would seem to bear out this theory. It is also perhaps significant that the quantity $P\cos\theta$ at the edge of the dome was approximately twice the value at the edge of the black ring. The theory, however, gives no account of the formation of separate 'spikes' in calm water, nor does it account for the increased maximum diameters of the dome and black ring when the water is at all rough.

Although there is experimental evidence for the assumption that water can sustain considerable tensions set up by shock waves (Wood & Lakey 1942), it is not known whether the value of this tensile strength is sufficiently constant over any considerable volume of water for continuous layers to be formed. Under static conditions, the negative pressure at which cavitation occurs varies enormously in the presence of local inhomogeneities or dissolved gases.

(*b*) *Multi-bubble theory*

Professor N. F. Mott in a private communication has drawn our attention to a theory to account for the phenomena of the 'crack' and of the formation of individual 'spikes' which does not assume a constant tensile strength for water.

When water is subjected to a large negative pressure, any small bubbles of gas present will expand rapidly to very many times their original size (e.g. with a shock wave of peak pressure—200 lb./sq.in., a bubble of radius greater than about 10^{-5} cm. will expand to a radius of the order of 1 cm.). It may also be shown that, as a result of surface tension, for each negative pressure there will be a minimum radius below which expansion will not take place.

This theory attributes the 'crack' to the expansion of large numbers of such bubbles as a result of the tension set up by the reflexion of the outgoing shock wave at the free surface. In the area of the black ring the peak pressure is smaller and the minimum diameter for a bubble which can expand is consequently greater; bubbles of larger radii will be rarer, and the expansion of those near the surface may account for the formation of individual 'spikes'.

The difficulties of this theory are, first, that if the distribution of bubble size in the water is continuous, it is difficult to account for the comparatively sharp outer edge of the black ring and, secondly, that as the expansion of the bubbles will depend only on the magnitude of the pressure and not on the direction of the wave-front, then P and not $P\cos\theta$, would be expected to be constant at the edges of the dome and black ring.

(*c*) *Accentuated-wave theory*

Sir Geoffrey Taylor in a private communication has put forward a theory to explain the breaking up of the water surface into drops by a shock wave. He has shown that when any water surface which is not perfectly smooth is subjected to an acceleration greater than that of gravity, the surface will become unstable provided that the wave-lengths of the irregularities obey certain conditions.

In the case of the large impulsive acceleration set up by a shock wave striking the surface of the water the general effect will be that any waves in the water will have their amplitude greatly increased, causing a roughening of the water as seen in the black ring. When this magnification is sufficiently great, the crests of the waves will break up into drops giving the white appearance of the dome. In this way the appearance of the dome and black ring can be accounted for apart from any questions of cavitation or the formation of bubbles.

The results already described in this report, for experiments in water that is at all rough, support this theory. Figure 7, plate 10, shows the waves in the black ring being enhanced, the crests of those nearest the charge breaking up into drops. The view of the dome in figure 5, plate 9, taken from the tower also shows that the structure of the dome itself is striated, as if it were conforming to the waves originally present in the water.

Although the effects predicted by Professor Taylor's theory certainly appear to take place, the theory gives no account of the 'crack' phenomena, the effects observed in very calm water, or the formation of separate spikes, and it would thus appear that one of the mechanisms assumed in the other two theories must also be involved.

(3) *Expansion of the bubble*

The motion of the gas-bubble produced by an underwater explosion has been very fully discussed by several workers and they have shown that if, to a first approximation, the bubble is regarded as spherical, then its radius a at any time t after the explosion is given by the differential equation

$$a\ddot{a}+\tfrac{3}{2}\dot{a}^2 = \frac{Pa_0^{3\gamma}-Qa^{3\gamma}}{\rho a^{3\gamma}}, \tag{9}$$

where a_0 is the radius of the bubble at $t = 0$ (taken as identical with the radius of the unexploded spherical charge), P is the initial pressure generated in the sphere of radius a_0 by the explosion, Q is the total pressure at the depth of the centre of the unexploded charge due to depth and atmosphere, ρ the density of water and γ the adiabatic constant of the gaseous explosion products. (It is assumed that the expansion of the gas is adiabatic.)

A first integral of this equation can be obtained using $a^2\dot{a}$ as an integrating factor, and if to get the equation in non-dimensional form we introduce the new variables $\alpha = a/a_0$ and $c_0 = \sqrt{(P/\rho)}$, we get

$$\frac{\dot{a}^2}{c_0^2} = \frac{2}{3(\gamma-1)}\left[\alpha^{-3}-\alpha^{-3\gamma}+\frac{Q(\gamma-1)}{P}(\alpha^{-3}-1)\right]. \tag{10}$$

We may take $\gamma = \frac{4}{3}$ as a reasonable approximation, in which case

$$\frac{\dot{a}^2}{c_0^2} = 2\left[\alpha^{-3}-\alpha^{-4}+\frac{Q}{3P}(\alpha^{-3}-1)\right] = \frac{2}{\alpha^4}\left[\alpha-1+\frac{Q}{3P}(\alpha-\alpha^4)\right]. \tag{11}$$

This enables us to calculate the *velocity* with which the bubble is expanding when its radius has reached a certain multiple α of the initial radius of the unexploded charge. This equation may be further integrated to determine the *time* taken for the bubble to reach this radius. This gives the integral:

$$\frac{c_0 t}{a_0} = \frac{1}{\sqrt{2}}\int_1^{\alpha}\frac{\alpha^2\,d\alpha}{\sqrt{(\alpha-1)}\left[1-\dfrac{\alpha+\alpha^2+\alpha^3}{\alpha_1^3}\right]^{\frac{1}{2}}}, \tag{12}$$

where $\alpha_1 = \sqrt[3]{(3P/Q)}$, we may expand the integrand, thus:

$$\frac{c_0 t}{a_0} = \frac{1}{\sqrt{2}} \int_1^\alpha \frac{\alpha^2 d\alpha}{\sqrt{(\alpha - 1)}} \left[1 + \frac{1}{2\alpha_1^3}(\alpha + \alpha^2 + \alpha^3) + \frac{3}{8\alpha_1^6}(\alpha + \alpha^2 + \alpha^3)^2 + \ldots \right], \tag{13}$$

and, since α_1^3 ($= 3P/Q$) is always very large, we may write as a good approximation

$$\frac{c_0 t}{a_0} = \frac{1}{\sqrt{2}} \int_1^\alpha \frac{\alpha^2}{\sqrt{(\alpha - 1)}} \left[1 + \frac{1}{2\alpha_1^3}(\alpha + \alpha^2 + \alpha^3) \right] d\alpha. \tag{14}$$

Writing $\alpha - 1 = z$, we can evaluate this integral, and obtain finally the result

$$\frac{c_0 t}{a_0} = \sqrt{(2z)} \left[1 + \frac{2z}{3} + \frac{1z^2}{5} + \frac{Q}{6P} \left(\frac{z^5}{11} + \frac{2z^4}{3} + \frac{15z^3}{7} + \frac{19z^2}{5} + 4z + 3 \right) \right]. \tag{15}$$

These results, for 1 and 10 lb. charges of 'P.E.', are tabulated in table 6, where $\dot{a}$ and t are evaluated for the bubble up to five times the initial radius of the charge. The calculations were made taking

$$P = 8{\cdot}19 \times 10^9 \text{ dynes/cm.}^2$$

and thus $\qquad c_0 = 9{\cdot}05 \times 10^4$ cm./sec.

It was found that the variation of Q due to varying the depth between 0·25 and 7 m. made no measurable difference to the values either of $\dot{a}$ or t in this range of the bubble's expansion. The calculations were therefore carried out using a value of $Q = 1{\cdot}2 \times 10^6$ dynes/cm.^{2}

TABLE 6

		1 lb. charges		10 lb. charges	
α ($= a/a_0$)	$\dot{a}$ (m./sec.)	a (cm.)	t (msec.)	a (cm.)	t (msec.)
1	0	4·11	0	8·85	0
1·1	334·5	4·5	0·022	9·74	0·47
1·2	397·5	4·9	0·033	10·62	0·071
1·3	414·8	5·3	0·043	11·5	0·092
1·4	413·0	5·75	0·053	12·4	0·114
1·5	402·2	6·2	0·063	13·3	0·135
2·0	320	8·2	0·120	17·7	0·258
2·5	250	10·3	0·193	22·1	0·415
3·0	211	12·3	0·285	26·6	0·613
3·5	165	14·4	0·397	31·0	0·796
4·0	138	16·4	0·533	35·4	1·148
4·5	118	18·5	0·699	39·8	1·504
5·0	102	20·6	0·881	44·25	1·895

Now in an infinite fluid it may be shown that the fluid velocity at any point at distance r from the centre of such an expanding bubble is given by

$$V = \frac{a^2}{r^2} \dot{a}, \tag{16}$$

where $\dot{a}$ is the rate of expansion of the bubble and a its radius at the instant considered. The fluid was not, of course, infinite in these experiments, so that the hydrodynamic system of perfectly radial motion will not strictly apply to these conditions.

The presence of the surface will bend the streamlines, as shown by Herring (1947). However, the formula above will still give an estimate of the velocities produced by the expanding bubble at points distant from it, provided the charge is a relatively deep one; and we can thus estimate the velocity effects that a deep charge would produce at the surface. It is easy to show in this way that even with the very high velocities associated with the expansion of the bubble in its early stages, the effects produced at the surface when the depth is greater than about 3 m. for a 10 lb. charge or 1·0 m. for a 1 lb. charge, are small in comparison with the effects already produced by the shock wave, while at depths greater still, the bubble produces no measurable effect at the surface during its first expansion.

With very shallow charges, however, formula (16) will not apply, for the radius of the bubble very rapidly approaches the depth of the charge, so that the bubble will break through the surface in its first expansion. Here, of course, the hydrodynamic assumptions on which the whole theory of the bubble's motion is based begin to break down; but they will still hold for the early stages of the expansion, so that the velocities calculated in table 6 will give estimates of those with which the bubble will break surface. Thus for a charge at a depth of a few times its original radius, we should expect the bubble to throw up a spout of water immediately with a velocity of the order of 300 m./sec.

Experiments carried out with very shallow charges show that this does happen. Figure 11, plate 11, shows the splash due to a 10 lb. charge at 50 cm., 10 msec. after detonation. The photograph was taken with a high-speed camera, the splash being illuminated with two flares above and behind it. It will be seen that a column of water quite different from the dome produced by deeper charges has been thrown up even at this early stage. The initial velocity of this column was measured and found to be 320 m./sec.

If the depth of the charge is increased the effects of the bubble will become smaller in magnitude and will take longer to appear, as the bubble will have expanded further before producing its full effect at the surface. Thus we should expect to find that the first effect of the explosion was the dome thrown up by the shock wave, but that this was very soon broken through by a spout of water produced by the bubble, the time taken for the break-through to occur being greater for increased depths. This phenomenon is in fact what is observed as the 'central spout', and experiments carried out with dyes show that the water thrown up in this spout came from immediately above the charge, which would seem to bear out the present theory of its formation.

Clearly there will be a certain depth of charge for which the velocity produced at the surface by the expansion of the bubble will be equal to that produced by the shock wave; for charges shallower than this, the initial velocity of the splash will be governed by the water thrown up by the bubble, and no persistent dome will be observed; for charges below this depth, the initial velocity of the splash will be governed by the water thrown up by the shock wave, and there will be a definite dome, broken through by a central column some time after its formation. It is possible to estimate this limiting depth by comparing the velocities given by equation (16) in conjunction with table 6 with those found experimentally, and given by equation

(1) (viz. $V = 66(W^{\frac{1}{3}}/D - 0{\cdot}1)$) at various depths D. It is found that this depth is about 1 m. for 10 lb. charges and 0·5 m. for 1 lb. charges.

Moreover, figure 18 shows that in a dye experiment with a 1 lb. charge at 0·5 m., the velocity of the top of the splash (due to the shock wave) was approximately equal to the velocity of the dyed water (due to the bubble).

There are three other features of the central column which it is possible to account for on this theory. First, as has already been mentioned, the water in the central column appears to be far less broken up than the water in the domes. Secondly, the central columns are of considerably smaller diameter than the domes, and thirdly, in several of the photographs (e.g. figure 1, plate 8) the water at the edges of the central column is seen to have a definite horizontal component of velocity. All these effects are to be expected if the central column is produced by a hydrodynamic motion of the water, resulting from the expansion of the bubble, whilst the domes are produced by the outgoing shock wave.

The authors' thanks are due to Professor G. I. Finch, F.R.S., for the hospitality of his laboratory, and for his interest and encouragement; to Lieut.-General E. C. Weston, C.B., Major R. J. B. Bolitho and Dr R. E. Slade for their support; and to their colleagues, Mr J. A. F. Gerrard, Mr F. A. Jacobs and Miss D. E. Spear for help in the work.

They would also like to thank Sir Geoffrey Taylor, F.R.S., and Professor N. F. Mott, F.R.S., for their interest and advice, and Messrs S. Barnett and J. Flynn of the U.S. Navy for help with the high-speed photography.

REFERENCES

Flower, W. D. 1928 *Proc. Phys. Soc.* **40**, 167.
Friedman, B. 1947 *Phys. Rev.* **72**, 177.
Herring, C. 1947 *Phys. Rev.* **72**, 170.
Hilliar, A. W. 1919 S.R.E. Dept., Admiralty, R.E. 142/19.
Hopkinson, B. 1914 *Phil. Trans.* A, **213**, 437.
Lamb, H. 1923 *Phil. Mag.* **45**, 257.
Penney, W. G. & Price, A. T. 1942 Ministry of Home Security, S.W. 27, R.C. 351.
Road Research Laboratory 1942–43 Undex 10, 16, 29, 38, 51 and 60. Admiralty.
Taylor, G. I. 1942 Ministry of Home Security, S.W. 19.
Temperley, H. N. V. 1943 Undex 64. Admiralty.
Wood, A. B. & Lakey, R. H. 1942 A.R.L./5/12.
Worthington, A. M. 1908 *A study of splashes*. London: Longman.

DESCRIPTION OF PLATES 8 TO 11

Note. The times shown beneath the photographs in figures 6, 7 and 10 refer to times after detonation.

PLATE 8

FIGURE 1. Splash from 10 lb. charge at 50 cm. depth.

FIGURE 2. Splash from 10 lb. charge at 400 cm. depth.

FIGURE 3. Splash from 1 lb. charge at 200 cm. depth.

PLATE 9

FIGURE 4. Dome from 10 lb. charge at 300 cm. depth in calm water, taken at 64 frames/sec.

FIGURE 5. Dome from 10 lb. charge at 500 cm. depth, 0·25 sec. after detonation. Photograph taken from top of tower.

FIGURE 6. Surface of water photographed with 12 in. lens for splash from 10 lb. charge at 700 cm. depth.

PLATE 10

FIGURE 7. Splash from 1 lb. charge at 700 cm. depth in rough water photographed with 12 in. lens.

FIGURE 8. Growth of 'crack' from 1 lb. charge at 200 cm. depth, taken at 2000 frames/sec.

FIGURE 9. Splashes from 1 lb. charges at 50 cm. depth with dye placed (*a*) on the surface of the water, and (*b*) directly above the charge respectively.

PLATE 11

FIGURE 10. Emergence of the bubble from a 1 lb. charge at 500 cm. depth.

FIGURE 11. High-speed photograph of 10 lb. charge at 50 cm. depth taken 0·01 sec. after detonation. Illumination provide by flare.

Experimental and theoretical studies of the behaviour of slow electrons in air. I

By L. H. G. HUXLEY, D.PHIL. AND A. A. ZAAZOU
Department of Electrical Engineering, University of Birmingham

(*Communicated by M. L. E. Oliphant, F.R.S.—Received* 25 *May* 1948—
Revised 6 *September* 1948)

This paper is an account of an experimental investigation of the motions of free electrons in air by the method developed by Townsend. An improved form of apparatus is described with the appropriate theory. The following parameters of the electronic motion were determined as functions of the ratio Z/p of the electric field strength Z to the gas pressure p: Townsend's energy factor k_T, the drift velocity W, the mean free path at unit pressure L and the mean proportion η of its energy lost in collisions with gas molecules. The experimental data are given in the form of tables and curves.

The drift velocity W is found by a new procedure based on the Hall effect and by comparing the velocities W so obtained with the direct measurements of W by Nielsen & Bradbury it is seen that the velocities of agitation are distributed approximately according to Druyvesteyn's law when Z/p exceeds 0·5. Bailey's factor G, which is of importance in ionospheric studies, is obtained from the experimental dependence of η on k_T.

Theoretical formulae are derived for k_T and W in terms of L, G and Z/p.

The theory of the new method for measuring W is given in an appendix.

1. INTRODUCTION

The work described in this paper was undertaken in the first instance to obtain measurements of the losses of energy of slow electrons in collisions with molecules of air, in order to use them in parallel investigations of the properties of the E-region of the ionosphere by means of the phenomenon of the interaction of radio waves.

The investigations have, however, proved to possess a wider interest than that originally envisaged. Although the properties of slow electrons in air had already been investigated by Townsend (1948) and Townsend & Tizard (1913) and by Bailey (1925), the agreement between their measurements is not always close, and there was need for a fresh investigation of electronic motion in air with modern vacuum techniques and an apparatus of improved design.

The new measurements show, in fact, that the previous results, while qualitatively correct, are in certain instances quantitatively untrustworthy and capable of leading to erroneous conclusions.

The aim of the present investigation was to obtain reliable measurements of the following parameters (Healey & Read 1941) of the electronic motion as functions of the ratio Z/p of the electric field strength Z to the gas pressure p, at room temperature (about 15° C):

(*a*) *Townsend's energy factor* k_T. This factor is the ratio

$$k_T = \frac{\text{mean energy of agitation of an electron}}{\text{mean energy of thermal agitation of a gas molecule}}$$

$$= Q/Q_t,$$

where $Q = \frac{1}{2}m\overline{U^2}$ and $Q_t = \frac{1}{2}M\overline{\Omega^2}$, and m and $\overline{U^2}$ are the mass and mean-square velocity of agitation of an electron, and M and $\overline{\Omega^2}$ the corresponding quantities for a gas molecule. k_T proves (at a fixed temperature) to be a function of the ratio Z/p. The root mean-square velocity $(\overline{U^2})^{\frac{1}{2}}$ is obtained immediately as

$$1{\cdot}15 \times 10^7 (k_T)^{\frac{1}{2}} \text{ cm.sec.}^{-1}.$$

(*b*) The drift velocity W of the centre of mass of a group of electrons through the gas, as a function of Z/p.

(*c*) The mean free path L of an electron at unit pressure = 1 mm. of mercury (1 Tor).

(*d*) The mean proportion η of its energy Q lost by an electron in a collision with a gas molecule; that is, the mean energy lost per collision is ηQ. From the values of η at the lower values of Z/p, we obtain Bailey's energy loss factor G which is of importance for the ionospheric investigations mentioned above.

2. MEASUREMENT OF THE TOWNSEND ENERGY FACTOR

2·1. *Principle of the method*

The method employed is that devised by Townsend (1948) for this purpose, in which k_T is obtained from a measurement of the ratio W/K of the drift velocity W to the diffusion coefficient K of an electron stream, but the measuring apparatus embodies important modifications to the original Townsend diffusion apparatus. The principle of the method is indicated in figure 1.

A stream of electrons, having already acquired a steady state of motion in a uniform electric field Z, enters the diffusion chamber, figure 1 (*a*), through an aperture A in the upper electrode and moves in the same uniform electric field Z towards the receiving electrode BC which is divided into a central portion B and an outer portion C. In the apparatus employed by Townsend and his school to study electronic motions, the receiving electrode was divided as shown in figure 1 (*b*), whereas the division in our apparatus is shown in figure 1 (*c*). In each case the spread of the electron stream by diffusion among the gas molecules is measured in terms of the ratio R of the current i_B to electrode B, to the total current (i_C+i_B) to the whole electrode.

In a co-ordinate system in which the electric field Z and the drift velocity W are parallel to Oz and the diffusion is isotropic, the differential equation for the electronic concentration n, in the steady state, within the diffusion chamber is (Healey & Reed 1941)

$$\nabla^2 n = \frac{W}{K}\frac{dn}{dz} = 2\lambda\frac{dn}{dz}, \tag{1}$$

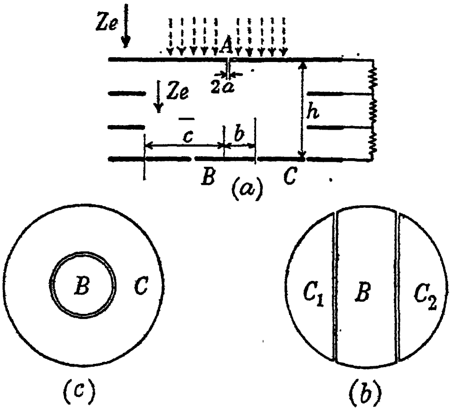


FIGURE 1. Diagram showing the diffusion chamber and two forms of receiving electrodes.

where $2\lambda = W/K$, and K is the coefficient of diffusion. Thus, in principle, when (1) has been solved in conformity with the boundary conditions the ratio of the currents, $R = i_B/(i_C+i_B)$, may be expressed as a function, $R(W/K, h, a, b, c)$ of the ratio W/K and the dimensions a, b, c and h of the diffusion chamber as shown in figure 1 (*a*).

2·2. *Formula for W/K in terms of Z and k_T* (Healey & Reed 1941)

Measurement of k_T

It is known that the correct formula for the drift velocity W of the centre of mass of a group of electrons moving through a gas in a constant and uniform electric field Z is

$$W = \frac{2}{3}\frac{Ze}{m}\overline{\left(\frac{l}{U}\right)}, \tag{2}$$

where e and m are respectively the electronic charge and mass, and l the mean free path of an electron whose velocity of agitation is U. In diatomic gases l is not strongly dependent on U and no detectable error is introduced in expressing (2) as

$$W = \frac{2}{3}\frac{e}{m}Zl\overline{(U^{-1})} = \frac{2}{3}\frac{e}{m}\left(\frac{Z}{p}\right)L\overline{(U^{-1})}, \tag{3}$$

where p is the pressure of the gas, L the mean free path at unit pressure (p = 1 mm. of mercury) and $\overline{(U^{-1})}$ the mean of the reciprocals of the velocity of agitation in the prevailing distribution of the velocities U along the free paths.

These formulae are derived on the hypotheses that the changes in the velocities U along a free path through the action of the force Ze, are small compared with $\overline{U}$, and that all directions of motion are equally probable after collision.

The corresponding formula for the diffusion coefficient K is

$$K = \tfrac{1}{3}l\overline{U}, \tag{4}$$

where p is the mean velocity of agitation. Thus from (3) and (4) it follows that

$$2\lambda = \frac{W}{K} = 2\frac{Ze}{m}\frac{(\overline{U^{-1}})}{(\overline{U})} = \frac{Ze}{\frac{1}{2}m\overline{U^2}}\left[\frac{(\overline{U^{-1}})(\overline{U^2})}{\overline{U}}\right]. \tag{5}$$

The factor in the square bracket is a dimensionless quantity whose value is determined by the law of distribution of the velocities U. Since $\overline{U^2}$ is the mean-square velocity it follows from § 1 that

$$\tfrac{1}{2}m\overline{U^2} = Q = k_T Q_t = k_T\frac{3}{2}\frac{R_0 T}{N}, \tag{6}$$

where R_0 is the gas constant, T the absolute temperature and N the number of molecules in a grammolecule of a gas. Thus (5) becomes

$$\frac{W}{K} = \frac{NeZ}{R_0 T}\frac{1}{k_1}, \tag{7}$$

where $$k_1 = Ak_T$$

and $$A = \frac{3}{2}\left[\frac{\overline{U}}{\overline{U^2}(\overline{U^{-1}})}\right]. \tag{8}$$

Since Ne is known accurately from electrolysis, and R_0 is the gas constant, it follows that for a temperature $T = (273+15)^\circ$ (7) becomes

$$W/K = 40{\cdot}3Z/k_1 = 40{\cdot}3Z/(Ak_T), \tag{9}$$

when Z is expressed in V/cm.

Thus, when the dimensions of the apparatus discussed in § 2·1 are given, the ratio R becomes $R(W/K, a, b, c, h)$, and it is possible to find W/K and therefore k_1 from the measured value of R.

To derive the energy factor k_T from k_1, the distribution of the velocities U must be known in order to determine A in (8). For instance, when the velocities U are distributed according to the law of Maxwell $A = 1$ and $k_1 = k_T$ (§ 3, table 1). When all the velocities U are the same, $A = \frac{3}{2}$.

It is known (Townsend 1948) that when k_T notably exceeds unity the velocities are not distributed according to Maxwell's law, but in many instances conform more closely (Healey & Reed 1941) to a law given by Druyvesteyn. In this event $k_1 = 1{\cdot}14k_T$.

2·3. *Description of apparatus*

In the form of diffusion apparatus employed by Townsend the aperture A (figure 1 (a)) was a slot whose half-width a entered into the expression for R which was obtained in the form of an infinite series. The computation of R as a function of Z/k_1 is therefore a laborious procedure which it is necessary to repeat whenever the dimensions of the apparatus are changed. For this reason, one of us has given an alternative solution (Huxley 1940) of equation (1) which is applicable to an apparatus in which the aperture A is a circular hole small enough to be regarded as a doublet source.

It was shown that the electron concentration n in the diffusion chamber is given closely by terms of the form

$$n = \text{const.}\, e^{\lambda z}(z/r)\frac{d}{dr}\left(\frac{e^{-\lambda r}}{r}\right),$$

and that the ratio R is
$$R = 1-\frac{h}{d}e^{-\lambda(d-h)}, \tag{10}$$

where r is the distance from the source, h the depth of the diffusion chamber,

$$\lambda = \frac{W}{2K} = 20\cdot15\frac{Z}{k_1} \quad (Z \text{ in V/cm.}),$$

$d^2 = h^2+b^2$, and b is the distance from the centre of the electrode B (figure 1 (c)) to the middle of the gap between B and C. In (10) R is the ratio of the current i_B to the total current entering through the hole A.

Expression (10) may be written

$$R = 1-\frac{1}{\sqrt{[1+(b/h)^2]}}\exp\left\{20\cdot15\frac{Zh}{k_1}\left(\sqrt{\left[1+\left(\frac{b}{h}\right)^2\right]}-1\right)\right\}. \tag{11}$$

Thus $R = f(Zh/k_1, b/h)$.

It is therefore a simple matter to calculate values of R for a range of values of Zh/k_1 when b/h is held fixed at a specified value. The relationship is then shown graphically by plotting R as ordinate against Zh/k_1 as abscissa. A family of such curves for a range of values of b/h is shown in figure 2.

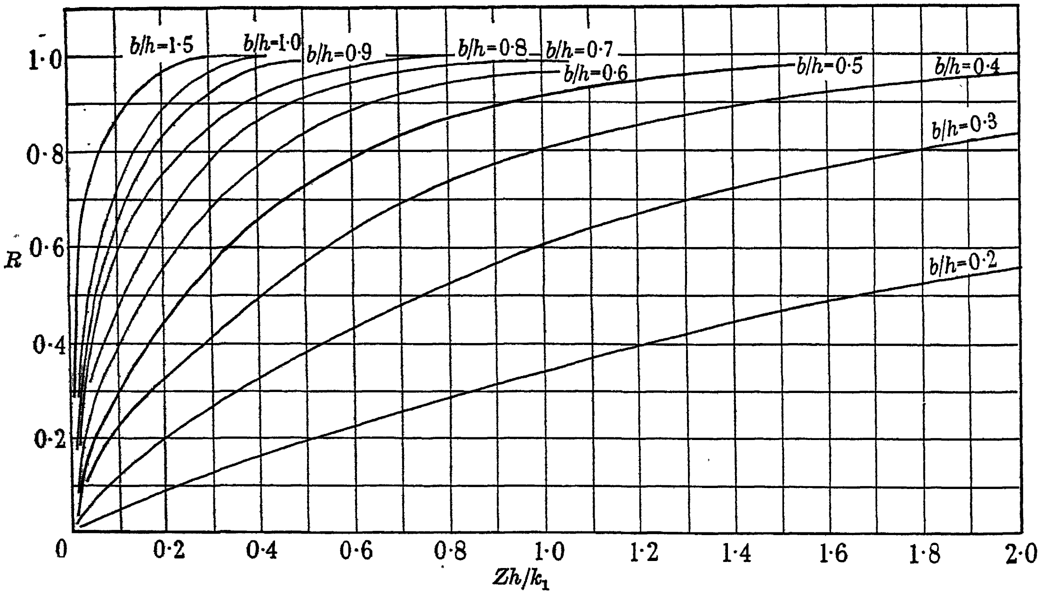


FIGURE 2. Curves showing the distribution ratio R as a function of (Zh/k_1) for a range of values of b/h. The significant curve in the present work is that for which $b/h = 0\cdot5$.

Each curve relates to one particular apparatus according to the value of b/h, which in these experiments was 0·5.

Figure 4 is a detailed drawing of the apparatus which was constructed of the same metal, nichrome, throughout to avoid contact potential differences. The aperture A (figure 3*a*) was a hole of diameter 1 mm., and the other dimensions of the diffusion chamber were:

$$\text{length } h = 1\text{ cm.},\quad \text{radius } b = 0{\cdot}5\text{ cm.},\quad \text{radius } c = 1{\cdot}5\text{ cm.}$$

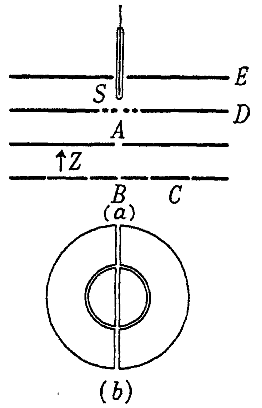


FIGURE 3. The diffusion apparatus—details of construction.

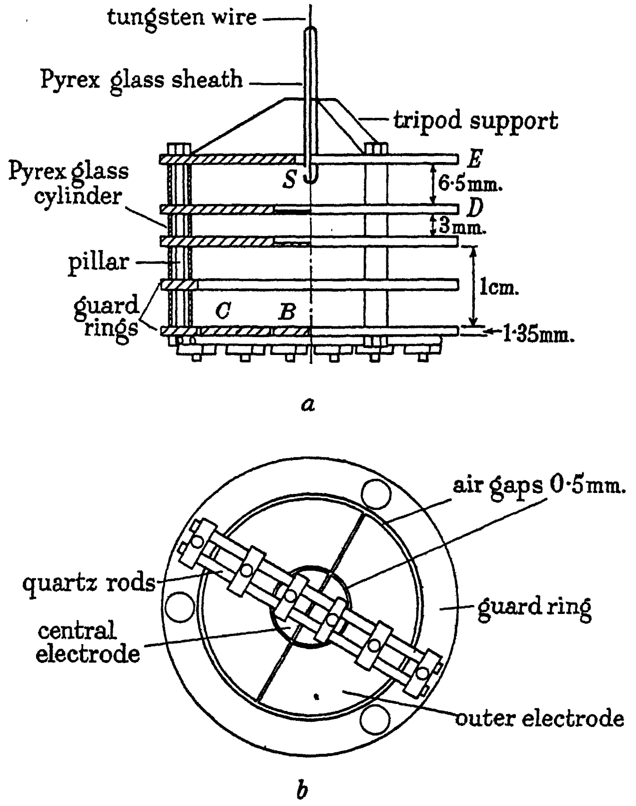


FIGURE 4. The diffusion apparatus—elevation in half-section. Plan of composite electrode.

The width of the annular gaps between the electrodes *B* and *C* and the electrode *C* and its guard ring (figure 3) was 0·5 mm. As shown in figure 4 the electrodes *B* and *C* are mounted on quartz rods attached to the outer guard ring which is earthed. It will be noted that the electrodes *B* and *C* are together divided by a diagonal cut (width 0·5 mm.), so that the receiving electrode comprises four distinct portions to which separate insulated leads are attached. Thus, by permutation of the external connexions the receiving electrode may be easily transformed from a system comprising a circular disk *B* surrounded by an annular electrode *C*, to a system of two semicircular electrodes. As described below, the latter system is used to measure the drift velocities.

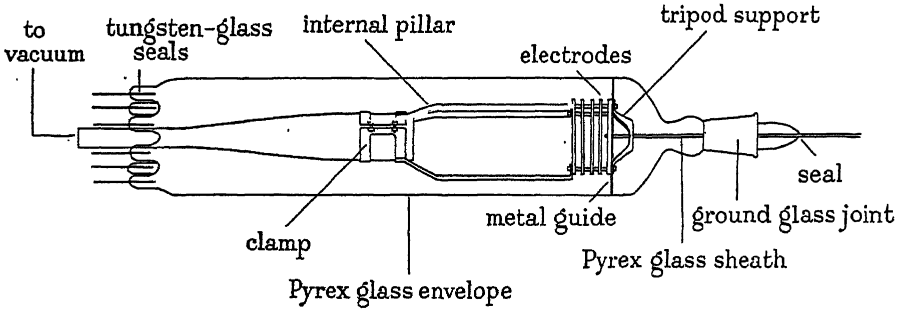


FIGURE 5. Apparatus in glass envelope.

The source of electrons was a brush discharge from a thin tungsten wire which is entirely enclosed within a pyrex glass sheath except for a hole about 2 mm. in diameter in the sheath several mm. from the end of the wire. The glass sheath serves to absorb ultra-violet radiation which would otherwise be scattered throughout the apparatus. The current in the brush discharge is of the order 10^{-7} to 10^{-6} A, and electrons emerging from the hole *S* are impelled by an electric field between *E* and *D* of about 50 V/cm. to the surface of *D* which is pierced by a cluster of small holes through which some of the electrons pass into the space between *D* and *A* where they acquire a steady state of motion in the same electric field *Z* as that prevailing between *A* and *BC* in the diffusion chamber. Some of the electrons enter the diffusion chamber through the hole *A* (1 mm. diameter) to form the stream whose lateral diffusion causes it to spread over *B* and a portion of *C*.

The brush discharge is fed from a stabilized d.c. power supply, through series resistors of 12 megohms and a diode valve whose filament emission serves to adjust the current. This forms a most convenient and stable source of electrons and appears to be greatly superior to the usual method of obtaining electrons by photoelectric emission.

An elementary estimate shows that a current of 10^{-6} A does not produce a detectable effect on the composition of the gas. The measurements were in fact uninfluenced by changing the current in the brush discharge. Although the current in the brush discharge is of the order 10^{-6} A, that passing through the hole *A* into the diffusion chamber is about 10^{-12} A. Consequently, the spread of the diffusing stream is uninfluenced by the space charge, which is negligible.

The whole electrode assembly is rigidly mounted on three symmetrically spaced pillars, each of which is a nichrome rod sheathed in a pyrex glass tube. Intermediate pyrex glass cylinders which fit over the inner glass sheath are carefully ground to correct lengths and serve as insulators and spacers between the electrodes. Lock nuts at the ends of the nichrome rods are used to clamp the whole electrode system into a rigid assembly. A mica sheet insulates the lock nuts and rods from the top plate *E* and from tripod support of the electron source.

The whole is enclosed within a pyrex glass envelope with an internal pillar from which the apparatus is supported, as shown in figure 5.

Connexions to all electrodes are taken through tungsten-to-glass seals at one end, except that to the brush discharge which is taken through a seal and a ground-glass joint at the other end as shown.

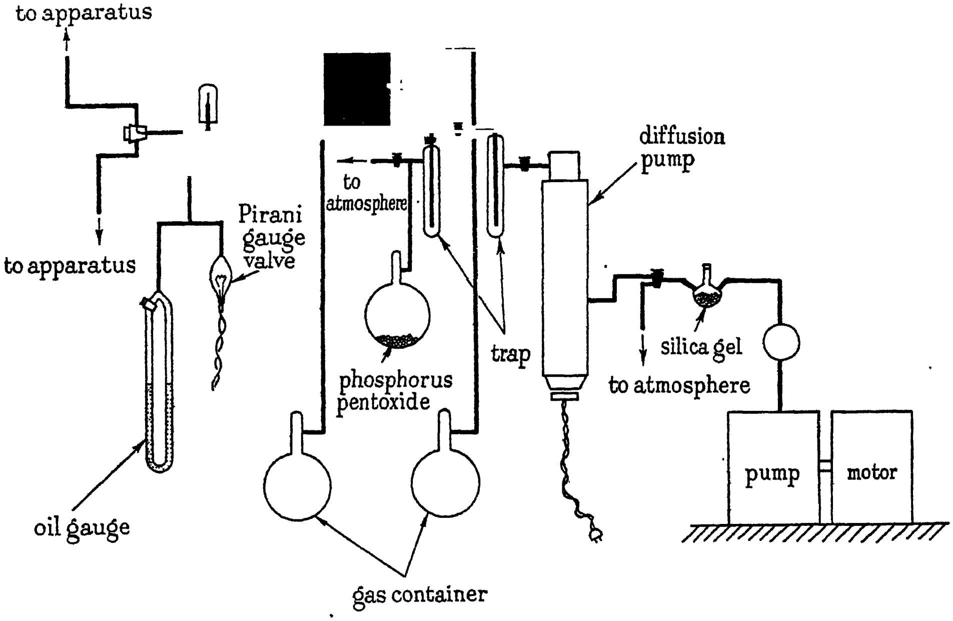


Figure 6. Vacuum installation.

2·4. *Vacuum system*

Apiezon L grease was used on all taps and joints, and gas pressures were measured on a manometer using apiezon B oil. An oil diffusion pump, Metrovac type 03 B with a Metrovac rotary backing pump together with a Pirani gauge, was used to evacuate the system to a pressure of 0·0001 mm. of mercury during the outgassing of the electrode assembly and to check that the system was completely airtight. Thus, mercury was employed nowhere. The vacuum installation is shown in figure 6.

To remove water vapour and carbon dioxide from the air, it was admitted to the apparatus after storage for a short time in the vacuum traps immersed in liquid air.

2·5. *Method of measuring small currents*

Since the currents to the electrodes *B* and *C* (figure 3) are of the order of 10^{-12} A it is necessary to employ electrometer techniques to measure them. Further, it is essential that the electrode B or C, the current to which is being measured, should not be allowed to change in potential during the measurements, in order to avoid distortion of the applied field Z in the diffusion chamber. A modification (Huxley 1936) of the Townsend's induction balance proves simple and convenient for this purpose. Figure 7 is a schematic diagram of connexions. To measure the current i_B, for instance, electrode C is earthed and electrode B is connected via an earthed screening tube to the grid of an electrometer tetrode valve. A substandard variable condenser is connected in shunt as shown.

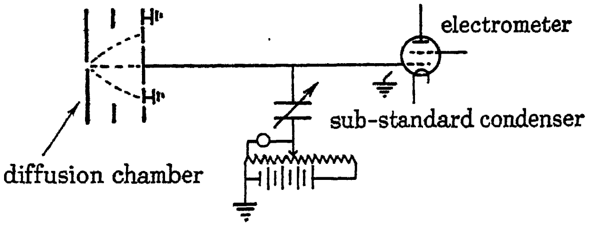


FIGURE 7. Schematic diagram of system measuring small currents.

This condenser comprises a good-quality variable radio condenser (0·0003 μF max.), mounted rigidly within a closed metal cylinder to which the movable plates are connected. Distrene insulation is used throughout, and the lead from the fixed plates is taken through a distrene plug in the wall of the cylinder. A pointer moves over a scale fixed to the top of the cylinder.

The system is calibrated by a null method against a set of standard cylindrical condensers.

The condenser stands on an insulating slab, and its case is held at a fixed potential of about 1 V as shown. As charge reaches the electrode the condenser capacitance is continuously altered to liberate a neutralizing charge at such a rate that the anode current of the valve indicates that its grid remains at earth potential. The current is therefore $i = v\Delta C/t$, where ΔC is the change in capacitance in time t (say 30 sec.) and v the potential of the case of the condenser.

The valve electrometer is mounted with its meters in a metal case with a front panel and forms a useful independent unit. By use of an arrangement in which the major portion of the anode current through the meter is 'backed-off', the sensitivity is greatly increased and it is possible to detect potential changes on the grid of the valve of the order of 1/100 V.

As the brush discharge was very steady the ratio $R = i_B/(i_B+i_C)$ was obtained simply by measuring i_B and (i_B+i_C) in turn, although a bridge method for measuring the instantaneous ratio i_B/i_C was also available.

Care was taken with earth connexions to avoid errors from contact potentials.

2·6. *Experimental results*

When the ratio $R = i_B/(i_B+i_C)$ has been determined for a given electric force Z and gas pressure p, as explained above, the constant k_1 is obtained from the appropriate curve in figure 2, which in these experiments was that designated $b/h = 0·5$.

Thus from R we find Zh/k_1 from the curve, and since Zh is the voltage across the diffusion chamber, k_1 is obtained immediately. It follows, from the discussion in § 2·2, that $k_T = k_1/A$, where A is a constant whose value is determined by the law of distribution of the velocities of agitation U (§ 3, table 1).

We shall in this paper consider two only of the possible laws of distribution, Maxwell's law for which $k_T = k_1$, and Druyvesteyn's law for which $k_T = k_1/1{\cdot}14$. When the values of k_1, obtained from measurements over a range of electric forces Z and gas pressures p, are plotted against the corresponding values of the ratio Z/p, the points fall near a single smooth curve. Thus k_1 is a function of Z/p. This result is also to be anticipated from the theory of the electronic motion (part II, § 1·3).

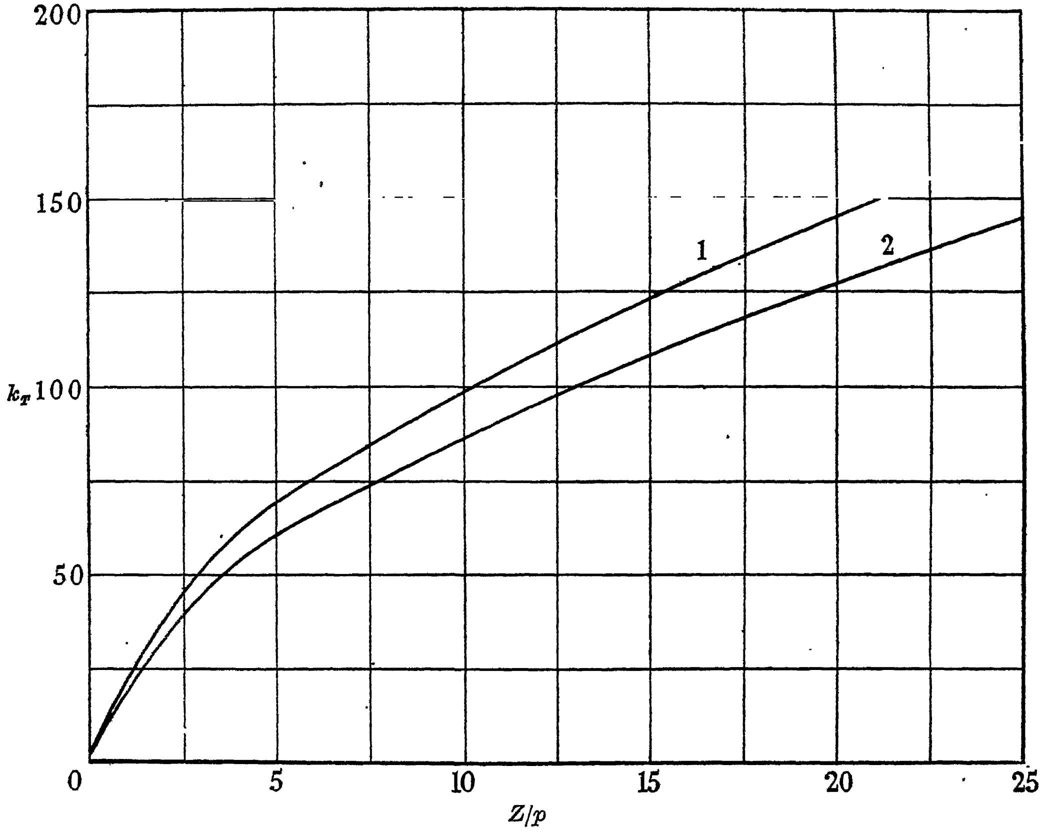


FIGURE 8. Curves giving Townsend's energy factor as a function of Z/p. Curve $1 = (k_T)_M$ for Maxwell's velocity distribution. Curve $2 = (k_T)_D$ for Druyvesteyn's velocity distribution.

The results are shown in graphical form in figure 8, in which the curve marked 1 represents the experimental values and that marked 2 is that derived from the upper curve by dividing its ordinates by 1·14. It therefore represents the values of k_T on the assumption of a Druyvesteyn distribution.

The corresponding values of the root mean square velocity of agitation in centimetres per second, are given by

$$\left.\begin{aligned}\sqrt{(\overline{U^2})} &= 1{\cdot}15\sqrt{k_T}\times 10^7 \text{ (Maxwell)}, \quad (k_1 = k_T)\\ \sqrt{(\overline{U^2})} &= 1{\cdot}08\sqrt{k_1}\times 10^7 \text{ (Druyvesteyn)}.\end{aligned}\right\} \qquad (12)$$

As shown later (§ 4·3), a sensitive test for the presence of ions in our experiments showed that with uncontaminated air there were so few that they escaped detection. It was therefore possible to use relatively large pressures of the order of 10 mm. of mercury in order to measure the parameters of the motion for small values of Z/p. In this way the field Z was never small enough to render contact potentials a serious source of error.

3. Ratios and products of velocity averages

The quantity A defined in equation (8) is an example of a dimensionless factor which appears as a combination of velocity averages of different kinds and whose value in consequence is determined only by the law of distribution of the velocities.

Since such factors will frequently appear in what follows, it is convenient to make a preliminary tabulation of the more useful examples, for the particular distribution laws of Maxwell and of Druyvesteyn.

We formulate these laws as follows:

Maxwell:
$$\frac{dN}{N}=\frac{4}{\alpha^3(\pi)^{\frac{1}{2}}}(\mathrm{e}^{-U^2/\alpha^2})\,U^2dU, \tag{13}$$

where α is the most probable velocity and dN/N is the proportion of a large group of electrons whose velocities exceed U but do not exceed $(U+dU)$ at any instant.

Druyvesteyn:
$$\frac{dN}{N}=\frac{4}{\alpha^3\Gamma(\frac{3}{4})}(\mathrm{e}^{-U^4/\alpha^4})\,U^2dU. \tag{14}$$

Here, α is not the most probable velocity.

Table 1. Dimensionless products and ratios of velocity averages

	Maxwell	Druyvesteyn
$\overline{U^2}(\overline{U^{-1}})/\overline{U}$	1·5	1·312
$\overline{U}(\overline{U^{-1}})$	$4/\pi = 1{\cdot}27$	1·18
$\overline{U}/(\overline{U^2})^{\frac{1}{2}}$	0·85	0·95
$(\overline{U^{-1}})\,\overline{(U^2)}^{\frac{1}{2}}$	1·38	1·24
$(\overline{U})^2/(\overline{U^2})$	$8/3\pi = 0{\cdot}847$	0·900
$(\overline{U^{-1}})^2/(\overline{U^{-2}})$	0·639	0·707
$(\overline{U^{-1}})^{\frac{3}{2}}/(\overline{U})^{\frac{1}{2}}\,(\overline{U^{-2}})$	0·564	0·65

The following formula proved useful in the preparation of the table:

$$\int_0^\infty \mathrm{e}^{-x^n}x^m dx=\frac{1}{n}\Gamma\left(\frac{m+1}{n}\right). \tag{15}$$

The following facts were also used:

$$\Gamma(\tfrac{1}{2})=(\pi)^{\frac{1}{2}};\quad \Gamma(\tfrac{1}{4})=3{\cdot}6216;\quad \Gamma(\tfrac{3}{4})=1{\cdot}2254;\quad \Gamma(\tfrac{5}{4})=0{\cdot}9064. \tag{16}$$

4. MEASUREMENT OF THE DRIFT VELOCITY W

4·1. *General considerations—Townsend's method*

In order to obtain information about the mean free paths of the electrons and the energy losses of electrons in collisions with molecules, it is necessary also to measure the drift velocities W.

The first measurements of W for electrons were made by Townsend (1948) by observing the deflexion of the stream of electrons in a magnetic field. By interchange of the external connexions, the receiving electrode (figure 1 (*b*)) of the diffusion apparatus, the electrodes B and C_2 could be connected to form a single electrode with C_1 behaving as a separate electrode (figure 9*a*). On applying a magnetic field parallel to the length of the centre electrode B the electron stream is deflected as shown (figure 9*b*). With particular values of the fields H and Z, the axis of the stream falls on the gap G, as is shown by the equality of the currents i_{C_1} and $(i_B+i_{C_2})$. An elementary application of the theory of the Hall effect gives the following formula for W:

$$W=\frac{Z}{H}\tan\theta=\frac{Z}{H}\left(\frac{b}{h}\right), \tag{17}$$

where Z in e.m.u./cm. and H in oersted are the values of and H necessary to make $i_{C_1}=(i_B+i_{C_2})$. A more careful analysis (Huxley 1937; Townsend 1937) shows that formula (17) is inaccurate.

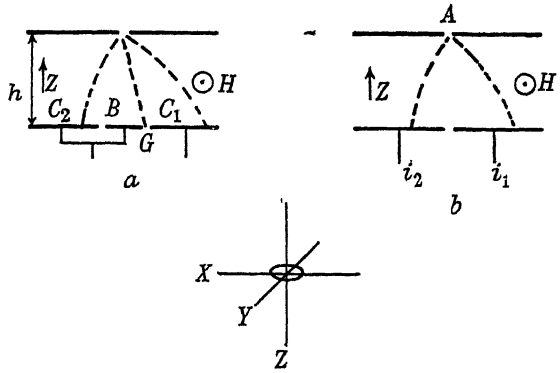


FIGURE 9. Hall effect in electron stream.

We define therefore an angle θ as follows:

$$\tan\theta=(W_x/W_z), \tag{18}$$

where W_x is the component of the drift velocity at right angles to the electric force Z and the magnetic field H and W_z is the component parallel to Z.

The following expressions have been derived (Huxley 1937) for the velocities W_x and W_z:

$$\left.\begin{aligned}W_x&=\frac{2}{3}\frac{Z}{H}\left\{\overline{\frac{\omega^2T^2}{(1+\omega^2T^2)}\left[\frac{1}{2}+\frac{\omega^2T^2}{1+\omega^2T^2}\right]}\right\},\\ W_z&=\frac{2}{3}\frac{Z}{H}\left\{\overline{\frac{\omega T}{(1+\omega^2T^2)}\left[1+\frac{\omega^2T^2}{1+\omega^2T^2}\right]}\right\},\end{aligned}\right\} \tag{19}$$

in which $\omega = He/m$, and $T = l/U$ = time to traverse the mean free parth l at speed U. The bar over the expressions in the curly brackets indicates that the mean values of these expressions averaged over all velocities U are to be taken.

In the diffusion experiments the gas pressures are such (1 to 6 mm. of mercury) that $\omega^2T^2 \ll 1$, consequently

$$\tan\theta = \frac{W_x}{W_z} = \frac{\omega\overline{T^2}}{2\overline{T}}. \tag{20}$$

Expression (20) may be written [equation (2)]

$$\tan\theta = \frac{1}{2}\frac{H}{Z}\frac{Ze}{m}\overline{T}(\overline{T^2}/(\overline{T})^2)$$

$$= \frac{3}{4}\frac{HW}{Z}[(\overline{U^{-2}})/(\overline{U^{-1}})^2], \tag{21}$$

or

$$W = \tfrac{4}{3}[(\overline{U^{-1}})^2/(\overline{U^{-2}})]\frac{Z}{H}\tan\theta. \tag{22}$$

Thus, from § 3, table 1, it follows that

$$W = 0{\cdot}85\frac{Z}{H}\tan\theta \quad \text{(Maxwell)}, \qquad W = 0{\cdot}943\frac{Z}{H}\tan\theta \quad \text{(Druyvesteyn)}. \tag{23}$$

In Townsend's measurements θ is determined by the method indicated in figure 8*a* and W is then given by formula (17).

4·2. *Method of Nielsen and Bradbury*

A totally different method for measuring W is the electrical shutter method originally used with ions by Van de Graaff and by Tyndall & Powell, and later adapted for use with electrons by Nielsen & Bradbury (1937).

Nielsen & Bradbury's results for air are discussed below.

4·3. *A new method*

In the method used in our experiments the quantity $W_x/W_z = \tan\theta$, as in Townsend's method, is obtained directly by experiment. The drift velocities W are then derived from (23). The method does not, however, depend upon the assumption that $W_x/W_z = \tan\theta = (b/h)$ (figure 9*a*) which is not strictly accurate (Huxley 1940), especially in a widely divergent stream.

The principle of the method and the experimental procedure (in contrast with the mathematical theory—see appendix) are extremely simple and may be understood by referring to figure 9*b*. As was stated in § 2·3, the receiving electrode shown in figure 3*b* can be transformed by rearrangement of the external leads, to a pair of separate semicircular electrodes. When the magnetic field H is zero, the currents i_1 and i_2 to the two semicircles are equal, but when H is applied parallel to the diagonal cut separating the electrodes, the currents i_1 and i_2 are no longer equal (figure 9*b*), and it is a simple matter to measure the ratio $R_1 = i_2/i_1$ of these currents by the method described in § 2·5.

It has fortunately proved possible to express the ratio R_1 as a function

$$R_1(W_x/W_z,\ Zh/k_1).$$

In figure 10 this functional relationship is shown graphically by a family of curves each of which represents R_1 as a function of Zh/k_1 for a specified constant value of W_x/W_z.*

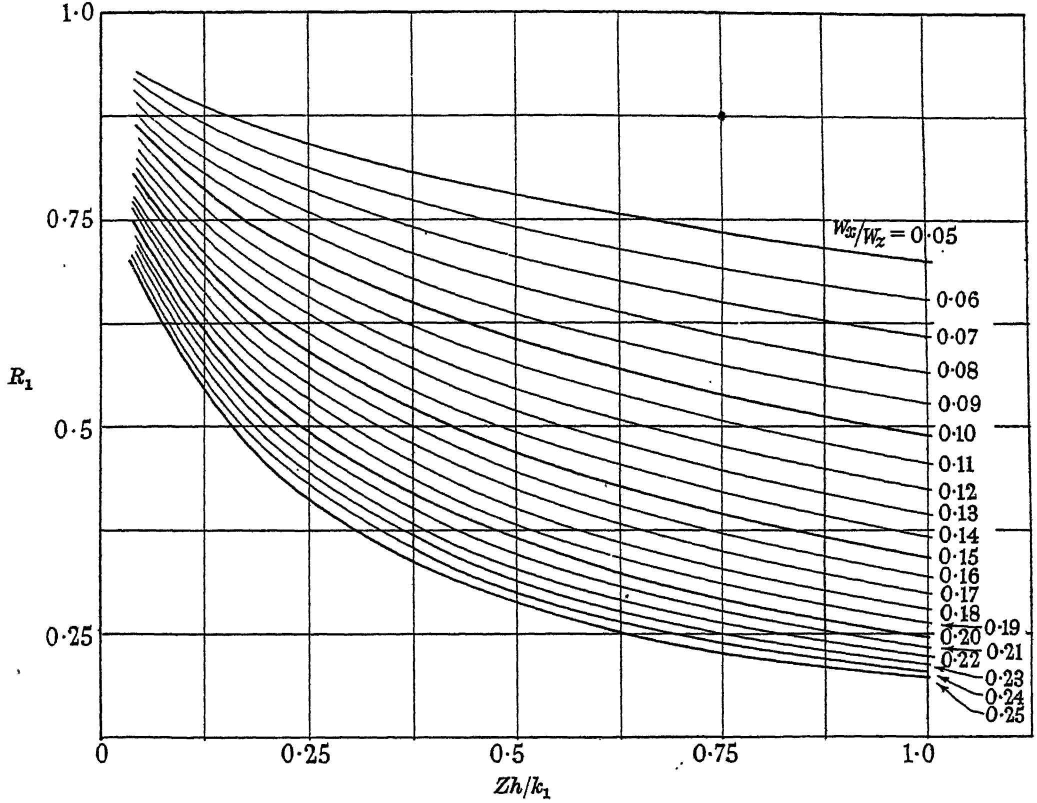


FIGURE 10. Hall effect ratio R_1 as a function of Zh/k_1 for a range of values of (W_x/W_z).

On the assumption that the stream of ions in the diffusion chamber is effectively that due to doublet source, it may be shown (see appendix) that

$$\frac{i_2}{i_1} = R_1 = \left(\frac{I_1 - I_2}{I_1 + I_2}\right), \tag{24}$$

where

$$\left.\begin{aligned} I_1 &= \frac{\pi}{2\lambda h}\exp\left[-\lambda h\{1-(\beta/\lambda)^2\}^{\frac{1}{2}}\right], \\ I_2 &= \sum_{n=0}^{\infty}(\beta/\lambda)^{(2n+1)}\frac{(\lambda h)^n}{1\,.\,3\,.\,5\,.\,.\,.\,.\,.\,(n-1)}K_n(\lambda h), \end{aligned}\right\} \tag{25}$$

---

* We are greatly indebted to Mr J. E. Goodwin of this department for computing these curves.

$$\left.\begin{aligned} &\text{where} && \lambda = \frac{(W_x^2+W_z^2)^{\frac{1}{2}}}{2K_H} \approx \frac{W}{2K} = 20{\cdot}15Zh/k_1 \quad [\text{eqn. (9)}] \\ &\text{and} && K_H = \left(\overline{\frac{U}{1+\omega^2T^2}}\right)\frac{l}{3}, \end{aligned}\right\} \tag{26}$$

$$\beta = \frac{W_x}{2K_H}, \quad (\beta/\lambda) = \frac{W_x}{W_z}, \tag{27}$$

and $K_n(\lambda h)$ is a modified Bessel function of the second kind and of order n.

Since, from eqn. (4), $\tan\theta = W_x/W_z \approx \frac{1}{2}\omega T$, and because in the experiments $\tan\theta < 0{\cdot}2$, $\omega^2T^2 \ll 1$, the approximations made in (26) and (27) are within the limits of experimental error. Thus, as stated, R_1 is a function $R_1(W_x/W_z, Zh/k_1)$.

Thus from the measured values of R_1 for known values of Z, p and therefore of Z/p, the value of $W_x/W_z = \tan\theta$ may be obtained from figure 10, since k_1 is already known as a function of Z/p (figure 8, curve 1).

The drift velocity W for the assumed law of distribution, either of Maxwell or of Druyvesteyn, is then obtained from (23).

The magnetic field was obtained from a pair of Helmholtz coils calibrated by means of a ballistic galvanometer and a small search coil. The measured and calculated values of H were in close agreement.

Table 2 shows how the same value of W is obtained for widely different values of H, when Z and p are given.

TABLE 2

$Z = 16$ V/cm. $p = 4$ mm. of mercury. $Z/p = 4$, $Zh/k_1 = 0{\cdot}242$

			$W \times 10^{-5}$ cm.sec.$^{-1}$	
H (oersted)	R_1	$W_x/W_z = \tan\theta$	Maxwell	Druyvesteyn
35·8	0·805	0·065	24·7	27·4
45·0	0·75	0·08	24·4	26·8
62·5	0·66	0·115	25·0	27·8
71·5	0·63	0·13	24·8	27·4

In addition to providing a satisfactory check of the theory, the fact that the same value of W is obtained over a range of values of H indicates that negative ions are present, if at all, to a negligible extent.

The drift velocities W were measured at several pressures p and forces Z over a wide range of values of Z/p. It was found that W was accurately a function of (Z/p) only, in accordance with (3).

The experimental results are shown in figure 11, in which W is shown as a function of Z/p. The curves marked W_M and W_D are obtained by using the same set of experimental values of $\tan\theta$ in the two formula (23), the one corresponding to the distribution law of Maxwell and the other to that of Druyvesteyn.

The experimental results of Nielsen & Bradbury (1937) obtained by the method of the electrical shutter are represented by the curve W_N. The curve W_N agrees well with the curve W_D, and it is therefore reasonable to suppose that the actual law of distribution of the velocities of agitation U is not very different from that of Druyvesteyn. The transition to the law of Maxwell which prevails when $Z = 0$ appears to occur in the range $0 < Z/p < 0{\cdot}5$.

Nevertheless, in what follows the values of numerical coefficients in formulae, appropriate to the distribution law of Maxwell as well as that of Druyvesteyn, are given, in order to indicate the influence of a change in the law of distribution on the accuracy of the results. Townsend's & Bailey's results for W when corrected on the hypothesis of a Druyvesteyn distribution are not very different from those of Nielsen & Bradbury and ourselves.

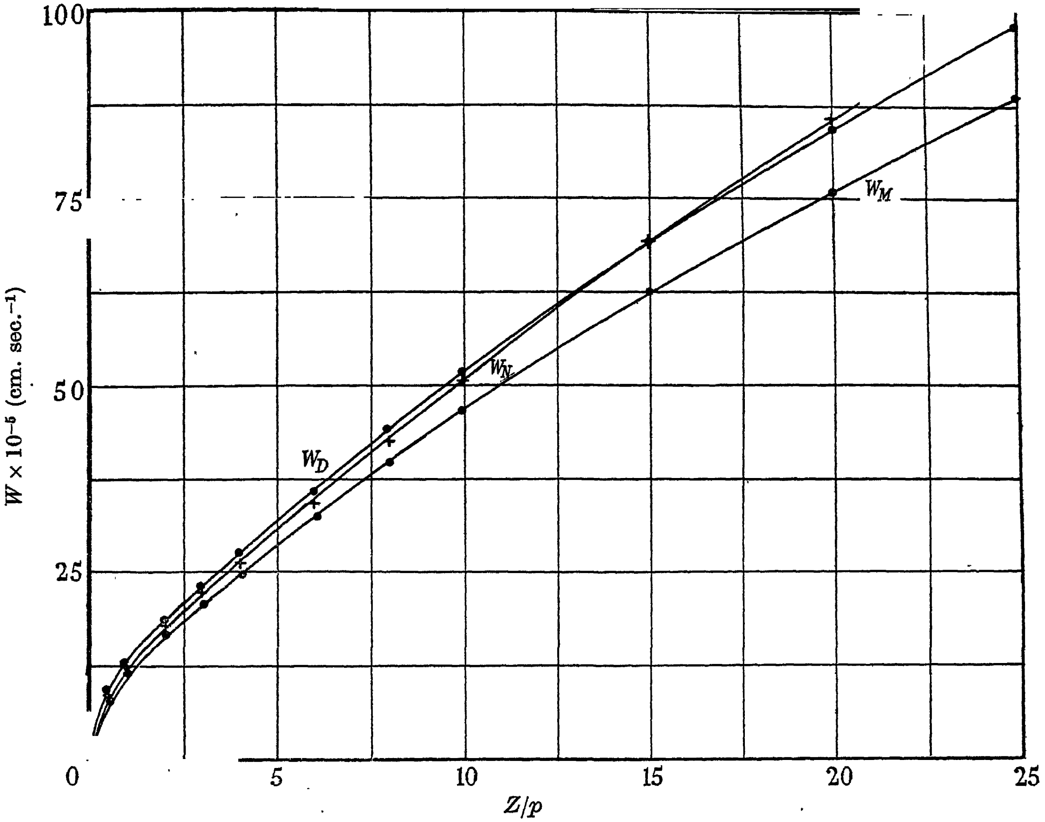


FIGURE 11. Drift velocities W as function of Z/p. W_N, Nielsen and Bradbury's measurements. W_D, from Hall effect assuming Druyvesteyn's distribution. W_M, from Hall effect assuming Maxwell's distribution.

5. CALCULATION OF THE MEAN FREE PATH L AT UNIT PRESSURES

When the drift velocity W and the experimental energy factor k_1 are known it is possible to deduce the mean free path L at unit pressure (1 mm. of mercury) from formula (3) from which it follows that

$$L = \frac{3}{2}\left(\frac{m}{e}\right)\frac{W}{(Z/p)}\frac{1}{(\overline{U^{-1}})}. \tag{28}$$

It is convenient to express L in terms of the experimental quantities and a factor determined by the distribution of the velocities U. If we consider W to be such an experimental quantity then we should use Nielsen & Bradbury's values for W and write (28) in the form

$$L = \frac{3}{2}\left(\frac{m}{e}\right)\frac{W}{(Z/p)}(\overline{U^2})^{\frac{1}{2}}[1/(\overline{U^{-1}})(\overline{U^2})^{\frac{1}{2}}]. \tag{29}$$

Since $$(\overline{U^2})^{\frac{1}{2}} = 1{\cdot}15\times 10^7(k_T)^{\frac{1}{2}}\,\text{cm.sec.}^{-1},$$

and, from (8) $$k_T = \frac{2}{3}\left[\frac{\overline{U^2}\,\overline{U^{-1}}}{\overline{U}}\right]k_1,$$

it follows that $$L = 7{\cdot}96\times 10^{-9}\left(\frac{W}{Z/p}\right)(k_1)^{\frac{1}{2}}[\overline{U}\,\overline{U^{-1}}]^{-\frac{1}{2}}, \tag{30}$$

in which Z is expressed in V/cm., p in mm. of mercury and W in cm.sec.$^{-1}$.

It follows from table 1, § 3, that with the two particular laws of Maxwell and Druyvesteyn,

$$\left.\begin{aligned} &\textit{Maxwell:} && L = 7{\cdot}05\times 10^{-9}\frac{W(k_1)^{\frac{1}{2}}}{(Z/p)}\,\text{cm.},\\ &\textit{Druyvesteyn:} && L = 7{\cdot}33\times 10^{-9}\frac{W(k_1)^{\frac{1}{2}}}{(Z/p)}\,\text{cm.}\end{aligned}\right\} \tag{31}$$

When W is derived from the magnetic deflexion of an electron stream it is then replaced in (30) by its value (22)

$$W = \tfrac{4}{3}\times 10^8[(\overline{U^{-1}})^2/(\overline{U^{-2}})]\frac{Z}{H}\tan\theta,$$

which becomes
$$\left.\begin{aligned} &\textit{Maxwell:} && W = 8{\cdot}5\times 10^7\frac{Z}{H}\tan\theta,\\ &\textit{Druyvesteyn:} && W = 9{\cdot}43\times 10^7\frac{Z}{H}\tan\theta,\end{aligned}\right\} \tag{32}$$

for substitution respectively in the first and second of equations (31). Z in (32) is expressed in V-cm.$^{-1}$. It is more satisfactory in this case to express L in terms of experimental quantities. It follows from (20) that

$$\tan\theta = \frac{\omega h}{2p}\left[\frac{\overline{U^{-2}}}{\overline{U^{-1}}}\right],$$

from which we deduce, using (8) and (12), that

$$L = 1{\cdot}06\left(\frac{Z}{H}\right)\frac{\tan\theta(k_1)^{\frac{1}{2}}}{(Z/p)}[(\overline{U^{-1}})^{\frac{3}{2}}/\{(\overline{U})^{\frac{1}{2}}\,\overline{U^{-2}}\}]\ \text{cm.}$$

Whence from § 3, table 1, last entry,

$$\left.\begin{aligned} &\textit{Maxwell:} && L = 0{\cdot}598\frac{Z}{H}\frac{\tan\theta}{(Z/p)}(k_1)^{\frac{1}{2}}.\\ &\textit{Druyvesteyn:} && L = 0{\cdot}689\frac{Z}{H}\frac{\tan\theta}{(Z/p)}(k_1)^{\frac{1}{2}}.\end{aligned}\right\} \tag{33}$$

The values of L derived from the second of formulae (33), using the measured values of $\dfrac{Z}{H}\dfrac{\tan\theta}{(Z/p)}$, are shown in figure 12 as functions of $\overline{U}$. Those obtained from Nielsen & Bradbury's values of W using formulae (30) are a few per cent less everywhere.

It is evident that L is a function of $\bar{U}$, and therefore of the energy of agitation. The agreement between the values of L obtained by two such different methods is noteworthy, and inspires confidence in the measurements of k_1 and W. On the assumption of a Druyvesteyn distribution $L \doteqdot 3{\cdot}7\times10^{-2}$ cm. $(5 < Z/p < 25)$. Figure 12 shows that L reaches its minimum value $L = 3{\cdot}7\times10^{-2}$ cm. at $U \approx 1{\cdot}5\times10^{8}$ cm.sec.$^{-1}$. The extrapolated value of L corresponding to thermal equilibrium at a temperature of 15° C ($\bar{U} = 1{\cdot}06\times10^{7}$ cm.sec.$^{-1}$) is $L = 5{\cdot}3\times10^{-2}$ cm. The value obtained by extrapolation of the L-$\bar{U}$ curve based on Nielsen & Bradbury's values of W give $L = 5\times10^{-2}$ cm. for $\bar{U} = 1{\cdot}06\times10^{7}$ cm.sec.$^{-1}$.

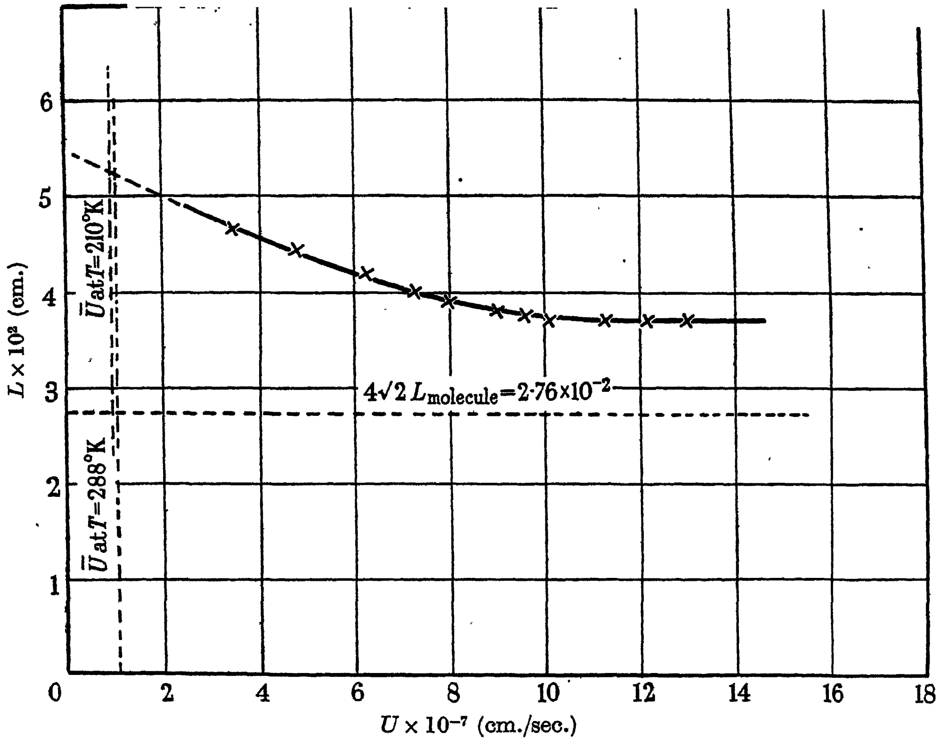


FIGURE 12. Mean free path L at unit pressure ($p = 1$ mm. of mercury) as a function of the mean velocity of agitation U. The horizontal dotted line is the value of $4\sqrt{2}L$ molecule. (L molecule from data given by Kennard, *Kinetic theory of gases*.)

6. THE PROPORTION η OF ITS ENERGY LOST ON THE AVERAGE BY AN ELECTRON IN A COLLISION WITH A MOLECULE

As an electron drifts at velocity W through the gas it receives power $w = ZeW$ from the electric field Z. In the steady state of motion this power is dissipated in collisions with the molecules, in aggregate, more energy being lost in some collisions than is gained in others.

Let $Q = \frac{1}{2}m\overline{U^2}$ be the mean energy of agitation of an electron in the steady state of motion, and suppose that on the average the overall loss of energy is ηQ per collision. It follows that

$$\left(\frac{\bar{U}}{l}\right)(\eta Q) = ZeW,$$

whence
$$\eta = 2\frac{ZeW}{m\overline{U^2}}\frac{l}{\overline{U}} = 3W^2/[(\overline{U^{-1}})(\overline{U})(\overline{U^2})]. \tag{34}$$

But
$$\overline{U^2} = (1{\cdot}15\times 10^7)^2 k_T = (1{\cdot}15\times 10^7)^2 \frac{2}{3}\left[\frac{\overline{U^2}\,\overline{U^1}}{\overline{U}}\right]k_1$$
$$= 8{\cdot}8\times 10^{13}\frac{(\overline{U^2})(\overline{U^{-1}})}{(\overline{U})}k_1, \tag{35}$$

whence
$$\eta = \frac{3{\cdot}41\times 10^{-14}}{(\overline{U^2})(\overline{U^{-1}})^2}\frac{W^2}{k_1}. \tag{36}$$

Thus from table 1, § 3, we have

$$\left.\begin{array}{ll}\textit{Maxwell}: & \eta = 1{\cdot}79\times 10^{-14}W^2/k_1.\\ \textit{Druyvesteyn}: & \eta = 1{\cdot}68\times 10^{-14}W^2/k_1.\end{array}\right\} \tag{37}$$

The values of η calculated from the second formula (37), using Nielsen & Bradbury's data for W, are shown in figure 13 as a function of k_T.

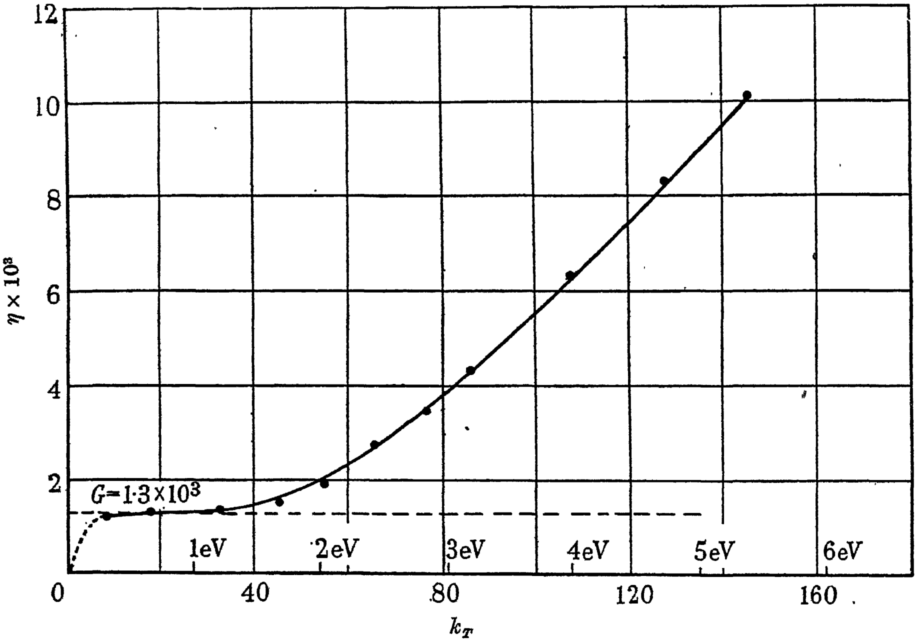


Figure 13. Mean proportion of energy η lost by an electron in collisions with gas molecules as a function of k_T. G, Bailey's energy loss factor.

Since η is zero when $k_T = 1$, it follows from the results that η increases with k_T to a constant value $\eta = 1{\cdot}3\times 10^{-3}$ and then rises continuously as k_T is further increased.

V. A. Bailey has supposed, in discussing the theory of the interaction of radio waves in the ionosphere, that η is given by a formula of the form

$$\eta = G(1-1/k_T), \tag{38}$$

which makes η vanish, as it should when $Z = 0$ when the electrons are in thermal equilibrium with the gas ($k_T = 1$).

Thus, according to (38), η should appear to preserve the constant value $\eta = G$ when $1/k_T$ is less than the experimental error.

This, in fact, occurs, and it may be assumed that when $Z/p < 2$ formula (38) represents the dependence of η on k_T, with

$$G = 1\cdot 3 \times 10^{-3} \quad \text{(Druyvesteyn's distribution)}.$$

7. FORMULAE FOR k_T AND W IN AIR IN TERMS OF Z/p

The measurements show that even when $Z/p < 2\cdot 5$ the mean free path L at unit pressure does not vary rapidly with the mean velocity $\overline{U}$, and that η is given by (38) with $G = 1\cdot 3 \times 10^{-3}$.

If Druyvesteyn's distribution is postulated then $k_1 = 1\cdot 14 k_T$, and equations (37) and (38) give $Z/p < 2\cdot 5$

$$1\cdot 47 \times 10^{-14} W^2/k_T = G(k_T - 1)/k_T, \tag{39}$$

while, from equation (31), we find

$$L = 7\cdot 85 \times 10^{-9} \frac{W}{(Z/p)} (k_T)^{\frac{1}{2}}. \tag{40}$$

From (39) and (40) we obtain formulae which express W and k_T in terms of L, G and (Z/p) only.

Thus, from (39) and (40),

$$W^2 = \frac{G(k_T - 1)}{1\cdot 47 \times 10^{-14}} = \left[\frac{L(Z/p)}{7\cdot 85 \times 10^{-9}}\right]^2 \frac{1}{k_T}, \tag{41}$$

in which Z is expressed in V.cm.$^{-1}$, whence

$$k_T(k_T - 1) = 240 \frac{L^2}{G} (Z/p)^2, \tag{42}$$

that is

$$k_T = \frac{1 + [1 + 960 L^2/G(Z/p)^2]^{\frac{1}{2}}}{2}, \tag{43}$$

also

$$W = \frac{L(Z/p)}{7\cdot 85 \times 10^{-9}} \frac{1}{(k_T)^{\frac{1}{2}}} = \frac{1\cdot 127 \times 10^8 L(Z/p)}{\left\{\frac{1 + [1 + 960 L^2 (Z/p)^2/G]^{\frac{1}{2}}}{2}\right\}^{\frac{1}{2}}}. \tag{44}$$

This formula for W is analogous to one given by K. T. Compton (see Cobine 1941), in which, however, the factor G is replaced by $2\cdot 66 m/M$, where m is the mass of the electron and M that of the gas molecule. Compton's formula did not represent the experimental results in diatomic gases very well. On the other hand, when $L = 4 \times 10^{-2}$ and $G = 1\cdot 3 \times 10^{-3}$ formulae (43) and (44) represent the dependence of k_T and W on Z/p fairly well provided $Z/p < 2\cdot 5$. With these values of L and G the formulae assume the special forms given in (45) and (48),

$$k_T = \frac{1 + \{1 + 1180 (Z/p)^2\}^{\frac{1}{2}}}{2}, \quad W = \frac{5 \times 10^7 (Z/p)}{\left\{\frac{1 + [1 + 1180 (Z/p)^2]^{\frac{1}{2}}}{2}\right\}^{\frac{1}{2}}}. \tag{45}$$

When $k_T \gg 1$, it follows from (42) that

$$k_T = \frac{15\cdot 4}{(G)^{\frac{1}{2}}} L(Z/p), \tag{46}$$

but that when $k_T \doteqdot 1$, then $$k_T \doteqdot 1 + \frac{240L^2(Z/p)^2}{G}. \tag{47}$$

A formula which may be used to estimate k_T when (Z/p) is small.

With Maxwell's distribution the factor 960 in (43) and (44) is replaced by 1440, and 1·27 in (44) by 1·42. The factor 240 in (42) becomes 360. When $1180(Z/p)^2 \gg 1$ formulae (45) become

$$k_T = 0\cdot 5 + 17\cdot 4Z/p, \quad W = 1\cdot 2 \times 10^7 (Z/p)^{\frac{1}{2}}, \tag{48}$$

with $Z/p < 2\cdot 5$.

TABLE 3. QUANTITIES DEPENDENT ON THE DISTRIBUTION OF THE VELOCITIES

	Maxwell	Druyvesteyn
§ 2·2		
$k_T = \frac{2}{3}\left[\frac{\overline{U^2}\,\overline{U^{-1}}}{\overline{U}}\right] k_1$	$k_T = k_1$	$k_T = 0\cdot 876 k_1$
§ 2·5 (12)		
$(\overline{U^2})^{\frac{1}{2}} = 1\cdot 15 \times 10^7 (k_T)^{\frac{1}{2}}$	$1\cdot 15 \times 10^7 (k_1)^{\frac{1}{2}}$	$1\cdot 08 \times 10^7 (k_1)^{\frac{1}{2}}$
$\overline{U} = \left[\frac{\overline{U}}{(\overline{U^2})^{\frac{1}{2}}}\right] (\overline{U^2})^{\frac{1}{2}}$	$1\cdot 06 \times 10^7 (k_1)^{\frac{1}{2}}$	$1\cdot 02 \times 10^7 (k_1)^{\frac{1}{2}}$
§ 5 (32)		
$W = \frac{4 \times 10^8}{3} \frac{Z}{H} \tan\theta\, [\overline{U^{-1}}]^2 [U^{-2}]^{-1}$	$8\cdot 5 \times 10^7 \frac{Z}{H} \tan\theta$	$9\cdot 43 \times 10^7 \frac{Z}{H} \tan\theta$
$\tan\theta = \frac{W_x}{W_z}$, Z in V-cm.$^{-1}$		
§ 5 (30)		
$L = 7\cdot 96 \times 10^{-9} \left(\frac{W}{Z/p}\right) (k_1)^{\frac{1}{2}} B$ cm.,	$7\cdot 05 \times 10^{-9} \frac{W(k_1)^{\frac{1}{2}}}{(Z/p)}$	$7\cdot 33 \times 10^{-9} \frac{W(k_1)^{\frac{1}{2}}}{(Z/p)}$
where $B = [\overline{U}\,\overline{U^{-1}})^{-\frac{1}{2}}$		
$L = 1\cdot 06 (k_1)^{\frac{1}{2}} \frac{Z}{H} \frac{\tan\theta}{(Z/p)} \left[\frac{(\overline{U^{-1}})^{\frac{3}{2}}}{(\overline{U})^{\frac{1}{2}} (\overline{U^{-2}})}\right]$	$L = 0\cdot 598 \frac{Z}{H} \frac{\tan\theta}{(Z/p)} (k_1)^{\frac{1}{2}}$	$L = 0\cdot 689 \frac{Z}{H} \frac{\tan\theta}{(Z/p)} (k_1)^{\frac{1}{2}}$
§ 6 (36)		
$\eta = \frac{3\cdot 41 \times 10^{-14} W^2/k_1}{[(\overline{U^2})\,(\overline{U^{-1}})^2]}$	$1\cdot 79 \times 10^{-14} W^2/k_1$	$1\cdot 68 \times 10^{-14} W^2/k_1$
In air for $Z/p < 2\cdot 5$		
$\eta = G(1 - 1/k_T)$	—	$G = 1\cdot 3 \times 10^{-3}$
When L and G are independent of Z/p		
$k_T(k_T - 1) = CL^2(Z/p)^2/G$	$C = 360$	$C = 240$
$k_T = \frac{1 + [1 + 4CL^2(Z/p)^2/G]^{\frac{1}{2}}}{2}$	$C = 360$	$C = 240$
$W = \frac{D \times 10^8 L(Z/p)}{\left\{\frac{1 + [1 + 4CL^2(Z/p)^2/G]^{\frac{1}{2}}}{2}\right\}^{\frac{1}{2}}}$	$D = 1\cdot 42$ $C = 360$	$D = 1\cdot 27$ $C = 240$

8. SUMMARY OF FORMULAE AND RESULTS

In this section the more important of the formulae that have been given are summarized for convenience. Cross-references are appended.

The Townsend energy factor k_T

$W/K = \dfrac{NeZ}{k_1 R_0 T}$, where k_1 is the apparent energy factor (7).

The Townsend energy factor is $k_T = k_1/A$,

where
$$A = \frac{3}{2}\left[\frac{\overline{U}}{\overline{U^2}(\overline{U^{-1}})}\right]. \tag{8}$$

Root mean square velocity of agitation

$$\begin{aligned}(\overline{U^2})^{\frac{1}{2}} &= 1{\cdot}15\times 10^7 (k_T)^{\frac{1}{2}}\\ &= 1{\cdot}15\times 10^7\left(\frac{k_1}{A}\right)^{\frac{1}{2}} = 0{\cdot}947\times 10^7\left[\frac{\overline{U^2}(\overline{U^{-1}})}{\overline{U}}\right]^{\frac{1}{2}}(k_1)^{\frac{1}{2}}\,\text{cm.sec.}^{-1}.\end{aligned}$$

Drift velocity W and diffusion coefficient K [§ 2·2]

$$W = \frac{2}{3}\frac{Zel}{m}(\overline{U^{-1}}) = \frac{2}{3}\frac{e}{m}(Z/p)(\overline{U^{-1}})\,L, \quad K = \tfrac{1}{3}l\overline{U}.$$

9. SUMMARY OF RESULTS

The experimental results are summarized for convenience in table 4, in which the measured quantities are expressed in terms of Z/p.

TABLE 4. THE MOTION OF ELECTRONS IN AIR

All derived magnitudes are obtained on the hypothesis of Druyvesteyn's law for the distribution of the velocities

Z/p	k_1	k_T	$\sqrt{(\overline{U^2})}\times 10^{-7}$	$\overline{U}\times 10^{-7}$	$L\times 10^2$	$W\times 10^{-5}$ (magnetic deflexion)	$W_N\times 10^{-5}$ (Nielsen & Bradbury)	$\eta\times 10^3$
0·5	11·0	9·65	3·58	3·48	4·65	9·43	9·0	1·24
1·0	20·5	18·0	4·88	4·62	4·45	13·2	12·5	1·28
2·0	37·5	32·9	6·62	6·24	4·20	18·8	17·5	1·37
3·0	52·5	45·5	7·78	7·35	4·00	23·5	22·0	1·560
4·0	62·0	54·5	8·5	8·03	3·90	27·4	26·0	1·830
6·0	75·0	65·7	9·35	8·84	3·80	35·8	34·5	2·670
8·0	87·0	76·3	10·1	9·5	3·75	43·8	42·5	3·480
10·0	98·0	86·0	10·7	10·1	3·70	51·4	50·0	4·28
15·0	124·0	108·5	12·0	11·35	3·70	68·5	68·5	6·36
20·0	146·0	128·0	13·0	12·3	3·70	84·0	85·5	8·38
25·0	166·0	145·5	13·9	13·1	3·70	98·0	100·0	10·10

L values are derived from W (magnetic deflexion).

We are greatly indebted to Mr H. J. Morris for his invaluable assistance in executing all the glass blowing for us, some of it of considerable difficulty, and also to Mr H. Kennett and Mr J. C. Ellams of the Departmental Workshop for constructing the apparatus with great skill.

APPENDIX

The theory of the new method for measuring the drift velocity W from the Hall effect

In § 4·3 a new method of measuring the drift velocity W in a steady field was described, but the theory of the method was not given. It is therefore presented in this section.

Consider the receiving electrode (figure 9*b*) to be bisected by a cut parallel to the axis Oy with the magnetic field H parallel to the cut and the electric field Z perpendicular to the plane of the electrode and parallel to Oz. Take the electron source (the hole A) as origin of co-ordinates. Let the components of W be W_x and W_z. The coefficient of diffusion parallel to H is unaffected by H, and has the value K, but in directions normal to H its value is

$$K_H = \frac{l}{3}\left(\frac{\overline{U}}{1+\omega^2T^2}\right). \tag{49}$$

The differential equation satisfied by the electron concentration n in the steady state ($dn/dt = 0$) is

$$K_H\left(\frac{\partial^2 n}{\partial x^2}+\frac{\partial^2 n}{\partial z^2}\right)+K\frac{\partial^2 n}{\partial y^2} = W_x\frac{\partial n}{\partial x}+W_z\frac{\partial n}{\partial z}. \tag{50}$$

Since we are not concerned with the y-co-ordinate we eliminate it as follows:

Let $q = \int_{-\infty}^{+\infty} n\,dy$, then since $\partial n/\partial y = 0$ at $y = \pm\infty$, we may integrate (50) with respect to y to obtain

$$\frac{\partial^2 q}{\partial x^2}+\frac{\partial^2 q}{\partial z^2} = 2\beta\frac{\partial q}{\partial x}+2\gamma\frac{\partial q}{\partial z},$$

where

$$2\beta = W_x/K_H, \quad 2\gamma = W_z/K_H.$$

Put $q = V\,e^{(\beta x+\gamma z)}$, then it follows that

$$\left.\begin{aligned}\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial z^2} &= \lambda^2 V,\\ \lambda^2 &= \beta^2+\gamma^2.\end{aligned}\right\} \tag{51}$$

where

A solution that satisfied the boundary conditions at the top plate is

$$V = A\frac{z}{r}K_1(\lambda r) = A\frac{z}{r}\frac{\partial}{\partial r}(e^{-\lambda r}),$$

where K_1 is a modified Bessel function of the second kind and order unity and A is a constant.

Thus

$$q = \frac{Az}{r}\,e^{(\beta x+\gamma z)}K_1(\lambda r).$$

The current i_1 to the right-hand electrode is proportional to

$$\left[\int_0^\infty q\,dx\right]_{z=h},$$

i.e.

$$i_1 \propto Ah\,e^{\gamma h}\int_0^\infty e^{\beta x}\frac{K_1[\lambda(x^2+h^2)^{\frac{1}{2}}]\,dx}{(x^2+h^2)^{\frac{1}{2}}}.$$

Similarly, the current to the other half of the electrode is

$$i_2 \propto Ah\, e^{\gamma h} \int_0^\infty \frac{e^{-\beta x}\, K_1[\lambda(x^2+h^2)^{\frac{1}{2}}]\, dx}{(x^2+h^2)^{\frac{1}{2}}},$$

and the ratio of the currents is

$$R_1 = \frac{i_2}{i_1} = \left[\int_0^\infty e^{-\beta x} \frac{K_1[\lambda(x^2+h^2)^{\frac{1}{2}}]\, dx}{(x^2+h^2)^{\frac{1}{2}}}\right] \bigg/ \left[\int_0^\infty e^{\beta x} \frac{K_1[\lambda(x^2+h^2)^{\frac{1}{2}}]\, dx}{(x^2+h^2)^{\frac{1}{2}}}\right]. \tag{52}$$

It is therefore necessary to evaluate the infinite integrals in (52).

We are greatly indebted to Professor G. N. Watson, F.R.S., for effecting the integrations as follows:

Put
$$\int_0^\infty \frac{e^{\pm\beta x}\, K_1[\lambda(x^2+h^2)^{\frac{1}{2}}]\, dx}{(x^2+h^2)^{\frac{1}{2}}} = I_1 \pm I_2 \quad (0<\beta<\lambda),$$

where
$$\begin{matrix} I_1 \\ I_2 \end{matrix} = \int_0^\infty \begin{matrix} \cosh \\ \sinh \end{matrix} (\beta x) \frac{K_1[\lambda(x^2+h^2)^{\frac{1}{2}}]}{(x^2+h^2)^{\frac{1}{2}}}\, dx.$$

By W.B.F.*, § 13·47 (2), we have

$$\int_0^\infty \frac{J_\mu(bt)\, K_\nu[a(t^2+z^2)^{\frac{1}{2}}]\, t^{\mu+1}\, dt}{(t^2+z^2)^{\frac{1}{2}}} = \frac{b^\mu}{a^\nu} \left[\frac{(a^2+b^2)^{\frac{1}{2}}}{z}\right]^{\nu-\mu-1} K_{(\nu-\mu-1)}\, [z(a^2+b^2)^{\frac{1}{2}}].$$

Put $\mu = -\frac{1}{2}$, $\nu = 1$ to find,

$$\int_0^\infty \left(\frac{2}{\pi bt}\right)^{\frac{1}{2}} \cos bt \frac{K_1[a(t^2+z^2)^{\frac{1}{2}}]\, t^{\frac{1}{2}}\, dt}{(t^2+z^2)^{\frac{1}{2}}}$$

$$= \frac{b^{-\frac{1}{2}}}{a} \left[\frac{(a^2+b^2)^{\frac{1}{2}}}{z}\right]^{\frac{1}{2}} K_{\frac{1}{2}}[z(a^2+b^2)^{\frac{1}{2}}]$$

$$= \frac{b^{-\frac{1}{2}}}{a} \left[\frac{(a^2+b^2)^{\frac{1}{4}}}{z^{\frac{1}{2}}}\right] \left[\frac{\pi}{2z(a^2+b^2)}\right]^{\frac{1}{2}} e^{-z(a^2+b^2)^{\frac{1}{2}}}.$$

That is
$$\int_0^\infty \frac{\cos bt\, K_1[a(t^2+z^2)^{\frac{1}{2}}]\, dt}{(t^2+z^2)^{\frac{1}{2}}} = \frac{\pi}{2az} e^{-z(a^2+b^2)^{\frac{1}{2}}}.$$

Put $b = i\beta$, $x = t$, $a = \lambda$, $z = h$ to obtain

$$I_1 = \frac{\pi}{2\lambda h} e^{-h(\lambda^2-\beta^2)^{\frac{1}{2}}} = \frac{\pi}{2\lambda h} e^{-\lambda h(1-\beta^2/\lambda^2{}^{\frac{1}{2}})}.$$

To evaluate I_2 consider

$$I_2 = \int_0^\infty \frac{\sinh(\beta t)\, K_1[a(t^2+z^2)^{\frac{1}{2}}]\, dt}{(t^2+z^2)^{\frac{1}{2}}}$$

$$= \sum_{n=0}^\infty \frac{\beta^{(2n+1)}}{(2n+1)!} \int_0^\infty \frac{t^{2n+1}\, K_1[a(t^2+z^2)^{\frac{1}{2}}]\, dt}{(t^2+z^2)^{\frac{1}{2}}} \quad (|\beta|<a)$$

$$= \sum_{n=0}^\infty \frac{\beta^{(2n+1)}}{(2n+1)!} \frac{(2z)^n\, n!}{a^{(n+1)}} K_{-n}(az) \quad \text{(by W.B.F., § 13·47 }(b)\text{)}$$

$$= \sum_{n=0}^\infty \frac{\beta^{(2n+1)}\, z^n a^{-(n+1)}}{1.3.5\ldots\ldots(2n+1)} K_n(az)$$

put $\lambda = a$, $z = h$

$$= \sum_{n=0}^\infty \frac{(\beta/\lambda)^{2n+1}\, (\lambda h)^n}{1.3.5\ldots\ldots(2n+1)} K_n(\lambda h).$$

---

* G. N. Watson (1944), *Theory of Bessel functions*, 2nd ed. Cambridge University Press.

Thus $$R_1 = \frac{i_2}{i_1} = \left(\frac{I_1 - I_2}{I_1 + I_2}\right).$$

But $$\beta/\lambda = W_x/W_z = \tan\theta, \quad 2\lambda = 2(\beta^2+\gamma^2)^{\frac{1}{2}} = W/K_H.$$

Since from (49) and 2 (7) and 2 (9), when $\tan\theta$ is small

$$\omega^2T^2 \ll 1, \quad \text{and} \quad K_H \to K; \qquad 2\lambda \to \frac{W}{K} = 40{\cdot}3Zh/k_1.$$

Thus $$R_1 = R_1(\tan\theta, Zh/k_1),$$

as stated in § 4·3. The curves in figure 10 were computed from the formulae for I_1 and I_2 given above.

REFERENCES

Bailey, V. A. 1925 *Phil. Mag.* (6), **50**, 825.

Cobine, J. D. 1941 *Gaseous conductors.* New York: McGraw Hill Book Co.

Healey, R. H. & Reed, J. W. 1941 *The behaviour of slow electrons in gases.* Amalgamated Wireless (Australasia) Ltd.

Huxley, L. G. H. 1936 *Phil. Mag.* (7), **22**, 1158.

Huxley, L. G. H. 1937 *Phil. Mah.* (7), **23**, 219.

Huxley, L. G. H. 1940 *Phil. Mag.* (7), **30**, 396.

Loeb, L. B. 1939 *Fundamental processes in electrical discharge in gases.* New York: John Wiley and Sons.

Nielsen, R. A. & Bradbury, N. 1937 *Phys. Rev.* **51**, 69.

Townsend, J. S. 1937 *Phil. Mag.* (7), **23**, 880.

Townsend, Sir John 1948 *Electrons in gases.* London: Hutchinson Sci. and Tech. Publs.

Townsend, J. S. & Tizard, H. T. 1913 *Proc. Roy. Soc.* A, **88**, 336.

Experimental and theoretical studies of the behaviour of slow electrons in air. II. Ionospheric and other applications

By L. G. H. Huxley, D.Phil.

Department of Electrical Engineering, University of Birmingham

(*Communicated by M. L. E. Oliphant, F.R.S.—Received* 25 *May* 1948—
Revised 6 *September* 1948)

The significance, for ionospheric and other studies, of the experimental data given in part I is considered. It is shown how the data obtained from rocket soundings may be combined with the measurements of wave interaction in the ionosphere and the laboratory data given in part I, to estimate the height at which wave interaction occurs and the collisional frequency at that height.

The following topics are also discussed: the theory of electronic motion in direct and alternating fields; the power communicated to a free electron in a gas by an alternating field; and the interaction of radio waves in the ionosphere.

1. Application to ionospheric studies

1·1. *The mean free path L*

It appears from figure 12, part I, that the mean free path L at a pressure of 1 mm. of mercury is about $5{\cdot}4\times10^{-2}$ cm. at the mean velocity $\overline{U} = 1{\cdot}05\times10^{7}$ cm.sec.$^{-1}$ corresponding to a temperature of 15° C (288° K), and that it has approximately the same value when $\overline{U} = 9\times10^{7}$ cm.sec.$^{-1}$ for a temperature of 210° K. If therefore the pressure p and temperature $T°$ K of the air in a region of the ionosphere are known, then the collisional frequency of an electron is $\nu = \frac{\overline{U}p}{L}\left(\frac{T}{288}\right)$, where T and L correspond to $\overline{U}$. The factor $(T/288)$ accounts for the thermal expansion or contraction of the air.

1·2. *The energy loss factor G (part I, § 6)*

According to the Bailey-Martyn theory (Healey & Reed 1941, p. 151) of wave interaction the dependence of the transferred modulation M upon the modulation frequency $(n/2\pi)$ of the interfering wave is given by the following formula (§ 4 below):

$$M = M_0/[1+(n/G\nu)^2]^{\frac{1}{2}}, \tag{1}$$

where M_0 is the value of M as $n\to0$, and G is the factor defined in part I, equation (38), whose value in air was found to be $G = 1{\cdot}3\times10^{-3}$; the collisional frequency ν is that of an electron at the seat of interaction in the E-region. Controlled observations (Huxley, Foster & Newton 1947, 1948; Ratcliffe & Shaw 1948) of wave interaction give the values of M_0 and $(G\nu)$.

Consider table 1 in which the first column contains a set of atmospheric heights ranging from 75 to 88 km. The second column contains the corresponding pressures p found from rocket soundings in America (Best, Havens & La Gow 1947). The corresponding collisional frequencies given by $\nu = \frac{\overline{U}p}{L}\left(\frac{2{\cdot}88}{2{\cdot}1}\right)$ are shown in the third column on the assumption that $T = 210°\,\mathrm{K}$, $L = 5{\cdot}4 \times 10^{-3}$ and $\overline{U} = 9 \times 10^6$ cm.sec.$^{-1}$, as indicated at the head of the table. The values of the product of the factor $G = 1{\cdot}3 \times 10^{-3}$ and the values of ν in the third column are shown in the fourth column.

TABLE 1

$T = 210°\,\mathrm{K}$. $\overline{U} = 9 \times 10^6$ cm.sec.$^{-1}$. $L = 5{\cdot}4 \times 10^{-2}$ cm. $G = 1{\cdot}3 \times 10^{-3}$

rocket data			$G\nu$ for air with molecular oxygen	estimated $G\nu$ for air with atomic oxygen
h (km.)	p (mm. Hg)	$\nu = \frac{\overline{U}p}{L}\left(\frac{2{\cdot}88}{2{\cdot}1}\right)$		
88	3×10^{-3}	7×10^5	910	505
85	5×10^{-3}	$1{\cdot}15 \times 10^6$	1500	830
82	7×10^{-3}	$1{\cdot}6 \times 10^6$	2100	1165
80	10^{-2}	$2{\cdot}3 \times 10^6$	3000	1670
78	$1{\cdot}3 \times 10^{-2}$	3×10^6	3900	2106
75	2×10^{-2}	$4{\cdot}6 \times 10^6$	6000	3333

In a series of test transmissions (Huxley *et al.* 1948) some twenty-three values of $G\nu$ were measured and it was found that $1100 < G\nu < 2300$, of which some 80 % of the values lay within the limits $1400 < G\nu < 2000$.

Since the collisional frequency varies rapidly with height this result suggests, with the assumption that the air at 82 km. is the same as clean laboratory air, that the region of wave interaction is centred on a height of 83 km. Although it is in fact improbable that at this height an appreciable proportion of the oxygen exists in the atomic form, it is of interest to estimate the height of the seat of interaction on the assumption that the oxygen is entirely atomic. In the first instance, if each molecule provides two atoms of roughly half its diameter, the collisional frequency is not greatly changed by the dissociation. In order to retain the pressures listed in the second column of table 1 the total number of particles, and therefore the collisional frequency, must, however, be reduced in the proportion of 5 to 6.

In normal air the losses of energy by electrons are expended in producing changes of rotational energy in the molecular dumbbells. In a monatomic gas, however, G is $2m/M$, where m and M are the electronic and atomic masses. Thus in atomic oxygen $G = 6{\cdot}8 \times 10^{-5}$, and may be neglected in comparison with $G = 1{\cdot}3 \times 10^{-3}$ for air with oxygen molecules. Thus the mean value of G for air comprising four parts of undissociated nitrogen and two parts of atomic oxygen is about $\frac{2}{3} \times 1{\cdot}3 \times 10^{-3}$. The value of $G\nu$ is therefore to be reduced by the factor $\frac{5}{6} \times \frac{2}{3} = \frac{5}{9}$ or approximately by one-half. The values of $G\nu$ in the fifth column of table 1 are $\frac{5}{9}$ of the corresponding values in the fourth column.

Since, as mentioned above, 80 % of the experimental values of $G\nu$ lie between 1400 and 2000, the corresponding height in this case is about 80 km. This shows that the height may be estimated with fair precision irrespective of an exact knowledge of the proportion of oxygen in the atomic form. As already stated, it may be safely assumed that at these heights the oxygen is molecular, so that if the values assumed for the pressures are correct, the seat of interaction lies at a mean height of about 83 km. Further experiments with rockets may of course require the data in the first two columns of table 1 to be revised.

A direct method of estimating the height of the seat of interaction from an examination of the phase of the impressed modulation has been described by Ratcliffe & Shaw (1948). This method at present does not lead to precise estimates of the height, but in the tests under discussion the measured heights ranged from 82 to 88 km. with about 60 % of them within the limits 82 km. $< h <$ 85 km. These heights are therefore consistent with the assumption of molecular oxygen and a height for the region of interaction of 83 km. as deduced above. Conversely, if the height h were reliably known then it would be possible to estimate the pressure p at that height from the radio measurements.*

1·3. *The significance of the Townsend parameter* (Z/p). *An alternative parameter* (w/p)

It has been shown by experiment that the energy factor k_T and the drift velocity W are functions of the ratio (Z/p) of the electric force Z to the pressure p. It is of interest to examine the theory of these functional relationships.

In the steady state of motion the balance between the power supplied by the field Z and that dissipated in collisions is given by equation (34), part I, which may be written

$$2ZeW = \eta(m\overline{U^2})\,\overline{U}p/L, \qquad (2$$

that is

$$\frac{\eta}{L^2}\left[\frac{\overline{U^2}\,\overline{U}}{\overline{U^{-1}}}\right] = \frac{3}{4}\left(\frac{m}{e}\right)^2 (Z/p)^2. \qquad (3)$$

It follows therefore that if Z and p are both increased or diminished by the same arbitrary factor so that Z/p is unaltered, then the product of terms on the left-hand side of (3) is unaffected. Since L and η are functions of $\overline{U}$, this condition can only be fulfilled in general if the individual averages, $\overline{U^2}$, $\overline{U}$ and $\overline{U^{-1}}$, depend only upon Z/p and not upon Z and p independently.

Thus Z/p determines both the mean energy of agitation $\frac{1}{2}m\overline{U^2}$ and the law of distribution of the velocities U, since the velocity averages $\overline{U^2}$, $\overline{U}$ and $\overline{U^{-1}}$ are also determined by it.

It follows that quantities such as

$$W = \frac{2}{3}\left(\frac{e}{m}\right)\left(\frac{Z}{p}\right)L(\overline{U^{-1}}) \quad \text{and} \quad k_1 = \frac{3}{2}\left[\frac{\overline{U}}{\overline{U^2}(\overline{U^{-1}})}\right]k_T$$

are functions of Z/p as is confirmed by experiment. Equation (1) may be written

$$\frac{w}{p} = \frac{ZeW}{p} = \eta\left(\frac{m\overline{U^2}}{2}\right)\frac{\overline{U}}{L}, \qquad (4)$$

where $w = ZeW$ is the mean power supplied to an electron by the field Z. Since the drift velocity W is a function $f(Z/p)$ which is found by experiment (part I, figure 10), it follows that

$$\frac{w}{p} = e(Z/p)\,f(Z/p) = \phi(Z/p),$$

or, inversely,

$$(Z/p) = \psi\left(\frac{w}{p}\right).$$

---

* The following formula gives p in terms of ν, T and $\overline{U}$: $p = 7{\cdot}6 \times 10^{-4}\,kT\nu/(A\overline{U})$ mm. of mercury, where k is Boltzmann's constant and A the molecular cross section in collisions with electrons. Since $Ah = 3{\cdot}13 \times 10^{-17}$, it follows that $A = 5{\cdot}8 \times 10^{-16}$ cm.^{2} when $h = 5{\cdot}4 \times 10^{-2}$ cm. Thus, with $\nu = 1{\cdot}2 \times 10^6$ and $T = 210°$, $p = 5 \times 10^{-3}$ mm. of mercury.

It also follows therefore that the quantities k_1, k_T, W, $\overline{U^2}$, $\overline{U}$ and the law of distribution of the velocities U are functions of (w/p), since they are functions of Z/p. This functional dependence of k_T upon the ratio w/p is of importance for the interpretation of the interaction of radio waves in the ionosphere where the power w is supplied to the electrons from the alternating field of the interfering wave. Expressions for w are given in § 3 below.

If it is possible to estimate the power w delivered by the wave to a free electron in the ionosphere, and if also the pressure p is known, then the energy factor k_T may be inferred from the results of the laboratory experiments assuming the temperatures to be roughly the same. The actual values of the energy factors k_T are maintained by the field of the interfering wave at values only a few per cent in excess of the value $k_T = 1$ corresponding to thermal equilibrium. We therefore require a formula expressing k_T as a function of (w/p) when $k_T \doteqdot 1$. When k_T is small (<2) we may assume the law of the velocity distribution to be that of Maxwell. It follows therefore, in accordance with the last sentence of § 7, part I, that

$$k_T(k_T-1) = 360\frac{L^2}{G}(Z/p)^2, \tag{5}$$

and
$$(w/p) = 10^8 e(Z/p)\,W = 1{\cdot}42\times10^{16}L(Z/p)^2 = 2{\cdot}27\times10^{-4}L(Z/p)^2, \tag{6}$$

since $e = 1{\cdot}6\times10^{-20}$ e.m.u. and Z is in V cm.$^{-1}$.

Thus, on eliminating $(Z/p)^2$ from (5) we find

$$(k_T-1)(k_T) = 1{\cdot}6\times10^6\left(\frac{L}{G}\right)\left(\frac{w}{p}\right). \tag{7}$$

L in cm., w in ergs sec.$^{-1}$, p in mm. of mercury.

The values $L = 5{\cdot}4\times10^{-2}$ and $G = 1{\cdot}3\times10^{-3}$ should be used in (7) since they are the experimental values of these quantities and are properties of the molecules and not of the law of distribution.

Thus, when $(k_T)\doteqdot 1$,

$$k_T \doteqdot 1 + 1{\cdot}6\times10^6\left(\frac{L}{G}\right)\left(\frac{w}{p}\right). \tag{8}$$

2. COMPONENTS OF THE DRIFT VELOCITY W IN A MAGNETIC FIELD AND THE DRIFT VELOCITY IN AN ALTERNATING ELECTRIC FIELD

The method of measuring the drift velocity W described in § 4·3, part I, depends on the formulae (19) for the components of the drift velocity in crossed electric and magnetic fields. It was found by experiment that if it is assumed that the velocities of agitation are distributed according to the law of Druyvesteyn, then the values of W obtained by this method agree closely with Nielsen & Bradbury's values, obtained by direct measurement over a very wide range of values of W. It may be concluded that the formula (20)

$$\tan\theta = \frac{\omega}{2}\frac{\overline{T^2}}{\overline{T}}, \tag{9}$$

which was derived from equation (19), part I, for the limiting case $\omega^2T^2 \ll 1$ has received experimental verification. The complete formulae (19), when $H = H_z$ and the electric force X is directed along Ox become

$$\left.\begin{aligned} W_y &= \frac{2}{3}\frac{Xe}{m}\left\{\frac{\omega T^2}{(1+\omega^2T^2)}\left[\frac{1}{2}+\frac{\omega^2T^2}{1+\omega^2T^2}\right]\right\}, \\ W_x &= \frac{2}{3}\frac{Xe}{m}\left\{\frac{T}{(1+\omega^2T^2)}\left[1+\frac{\omega^2T^2}{1+\omega^2T^2}\right]\right\}, \end{aligned}\right\} \tag{10}$$

where e is considered positive, $\omega = He/m$, $T = l/U = 1/\nu$.

On the other hand, it is known from electrodynamics that when the electron moves in a vacuum ($\omega T \to \infty$) then its mean velocity at right angles to the magnetic field is X/H. Equations (10) are in agreement with this fact since, when $\omega T \to \infty$,

$$W_y \to X/H, \quad W_x \to 0.$$

Thus equations (10) are correct for the two limiting cases $\omega T \to 0$ and $\omega T \to \infty$, and it may be inferred, pending further experimental tests, that they are generally valid.

We consider next the expression for the drift velocity in an alternating electric field, first without and then with an applied magnetic field. Let the electric field be

$$Z = Z_0 \cos p_0 t.$$

It was found (Huxley 1937) that the drift velocity W is

$$W = A\cos(p_0t) + B\sin(p_0t),$$

where

$$\left.\begin{aligned} A &= \frac{2}{3}\frac{Z_0e}{m}\left\{\frac{T}{(1+p_0^2T^2)}\left[1+\frac{p_0^2T^2}{1+p_0^2T^2}\right]\right\}, \\ B &= \frac{2}{3}\frac{Z_0e}{m}\left\{\frac{p_0T^2}{(1+p_0^2T^2)}\left[\frac{1}{2}+\frac{p_0^2T^2}{1+p_0^2T^2}\right]\right\}, \end{aligned}\right\} \tag{11}$$

in which $T = l/U$ and e is considered to be positive.

When (10) and (11) are compared, it will be noted that B and A are the same functions of p_0 as W_y and W_x are of ω. Further, equations (11) are correct in the limiting cases $p_0T \to 0$, and $p_0T \to \infty$ for which W becomes respectively,

$$W = \frac{2}{3}\frac{Z_0e}{m}\left(\frac{l}{U}\right) \quad \text{and} \quad W = \frac{Z_0e}{mp_0}\sin(p_0t).$$

The correspondence of (11) and (10) is not a coincidence, as it may be inferred from general considerations. If this be assumed then an experimental verification of (10) may be considered also to include (11). It may also be inferred (Huxley 1937, 1940) that (10) and (11) are each special cases of more general formulae for the components of W in an alternating electric field in the presence of a steady magnetic field. Suppose the electric field to be that of an elliptically polarized wave and that a magnetic field H is directed along Oz. Let the components of the electric field be

$$E_x = X_0 e^{jp_0t}, \quad E_y = Y_0 e^{j(p_0t+\alpha)}, \quad E_z = Z_0 e^{j(p_0t+\beta)}.$$

Then it is found that

$$\left.\begin{aligned} W_x &= \frac{e}{3m}\{[S-js]\,E_x+[d+jD]\,E_y\},\\ W_y &= \frac{e}{3m}\{[S-js]\,E_y-[d+jD]\,E_x\},\\ W_z &= \frac{2e}{3m}[G-jg]\,E_z, \end{aligned}\right\} \tag{12}$$

where

$$\left.\begin{aligned} S &= F(p_0+\omega)+F(p_0-\omega),\\ s &= f(p_0+\omega)+f(p_0-\omega),\\ D &= F(p_0+\omega)-F(p_0-\omega),\\ d &= f(p_0+\omega)-f(p_0-\omega), \end{aligned}\right.$$

and

$$\left.\begin{aligned} F(p_0\pm\omega) &= \overline{\left\{\frac{T}{1+(p_0\pm\omega)^2T^2}\left[1+\frac{(p_0\pm\omega)^2T^2}{1+(p_0\pm\omega)^2T^2}\right]\right\}},\\ f(p_0\pm\omega) &= \overline{\left\{\frac{(p_0\pm\omega)\,T^2}{1+(p_0\pm\omega)^2T^2}\left[\frac{1}{2}+\frac{(p_0\pm\omega)^2T^2}{1+(p_0\pm\omega)^2T^2}\right]\right\}}. \end{aligned}\right\} \tag{13}$$

The quantities G and g are $G = [F]_{\omega=0}, \quad g = [f]_{\omega=0}.$ (14)

The theory of wave propagation in the ionosphere, based on equations (12), has been given elsewhere (Huxley 1940).

3. THE MEAN POWER COMMUNICATED BY A HIGH-FREQUENCY FIELD TO AN ELECTRON IN A GAS

The mean power w given to an electron by the alternating field is

$$w = \frac{e}{2}Re[E_xW_x^*+E_yW_y^*+E_zW_z^*], \tag{15}$$

in which W_x^*, W_y^* and W_z^* are the complex conjugates of the components of W that appear in (12).

Thus $$w = \frac{e^2}{6m}[S(X_0^2+Y_0^2)+2GZ_0^2-2X_0Y_0D\sin\alpha]\ \text{ergs/sec.} \tag{16}$$

In the particular case of circularly polarized waves $X_0 = Y_0$, $\sin\alpha = \pm 1$. Suppose $Z_0 = 0$, then

$$\left.\begin{aligned} w &= \frac{2^2}{3m}X_0^2F+ && (\sin\alpha = +1)\\ \text{or}\quad w &= \frac{2^2}{3m}X_0^2F- && (\sin\alpha = -1), \end{aligned}\right\} \tag{17}$$

in which F^+ and F^- are defined in (13) and $\omega = He/m$ is to be given positive values in the formulae when H is directed along $+Oz$ and the charge e is positive. When H is reversed or e is negative then ω is negative.

If the wave is plane polarized with $E_y = 0$ and $\beta = 0$, then

$$w = \frac{e^2}{3m}\left[\frac{SX_0^2}{2} + GZ_0^2\right]. \tag{18}$$

When $H = 0$ this becomes

$$w = \frac{Ge^2}{3m}(X_0^2 + Z_0^2) = \frac{Ge^2}{3m}\,|\,E^2\,|. \tag{19}$$

In these formulae for w, e, X_0, Y_0 and Z_0 are expressed in e.m.u.

4. INTERACTION OF RADIO WAVES IN THE IONOSPHERE

The principles of electronic motion in gases which have been discussed in this paper find immediate application in the theory of the interaction of radio waves first given by Bailey and Martyn and later extended and reformulated by Bailey. We therefore give, in outline, a simplified formulation (Huxley *et al.* 1947) of the theory.

When the radiation from a powerful transmitter impinges on the E-region of the ionosphere its field communicates to each free electron, power w which is given by formula (16). The mean kinetic energy Q of the electrons is therefore increased to a value

$$Q = k_T Q_t, \tag{20}$$

where Q_t is the value of Q when w is zero. k_T is given by (8).

Suppose all the components of the field of the radiation to be modulated at a low frequency $n/2\pi$ so that the amplitude of each becomes multiplied by a factor $(1 + D\cos nt)$.

The power w now becomes a function of the time

$$w(t) = w_0[1 + D\cos nt]^2 = w_0\left[\left(1 + \frac{D^2}{2}\right) + 2D\cos nt + \frac{D^2}{2}\cos nt\right], \tag{21}$$

where w_0 is the value of w when $D = 0$, i.e. it is the contribution of the carrier wave to w.

The energy of agitation Q of the electrons also becomes a function of time and oscillates about a mean value $\bar{Q}$.

Equation (2) for the energy balance is now replaced by

$$\frac{dQ}{dt} + \eta Q\nu = w(t). \tag{22}$$

Since $(k_T - 1)$ is here a small quantity we may

(*a*) Put $\quad \eta = G\left(1 - \frac{1}{k_T}\right) = \frac{G}{Q}(Q - Q_t)$ (part I, equation (38)).

(*b*) Replace the frequency of collision $\nu = \overline{U}/l$, which oscillates with Q by its mean value $\bar{\nu}$ throughout the cycle corresponding to the mean energy $\bar{Q}$.

Thus from (21) and (22) we have

$$\frac{dQ}{dt} + G\bar{\nu}(Q - Q_t) = w_0\left[\left(1 + \frac{D^2}{2}\right) + 2D\cos nt + \frac{D^2}{2}\cos 2nt\right]. \tag{23}$$

The energy Q fluctuates about a mean value $\bar{Q}$ at which it is held by the mean power $w_0(1+D^2/2)$. The mean energy $\bar{Q}$ is therefore given by

$$G\bar{\nu}(\bar{Q}-Q_t) = w_0(1+D^2/2). \tag{24}$$

Subtract (24) from (23) to obtain

$$\frac{dQ}{dt}+(G\bar{\nu})(Q-\bar{Q}) = 2w_0\left[D\cos nt+\frac{D^2}{4}\cos 2nt\right]. \tag{25}$$

Thus, the fluctuation of Q about $\bar{Q}$ is given by

$$Q = \bar{Q}[1+2m\cos(nt-\phi)+2m'\cos(2nt-\phi')], \tag{26}$$

where

$$\left.\begin{aligned} m &= m_0/[1+(n/G\bar{\nu})^2]^{\frac{1}{2}}, \\ m_0 &= w_0D/[\bar{Q}(G\bar{\nu})], \\ m' &= m_0'\bigg/\left[1+\left(\frac{2n}{G\bar{\nu}}\right)^2\right]^{\frac{1}{2}}, \\ m_0' &= Dm_0/4, \\ \tan\phi &= \tfrac{1}{2}\tan\phi' = (n/G\bar{\nu}). \end{aligned}\right\} \tag{27}$$

The instantaneous and mean frequencies of collision ν and $\bar{\nu}$ are related as follows, $(\nu/\bar{\nu}) = (Q/\bar{Q})^{\frac{1}{2}}$, whence from (26)

$$\left.\begin{aligned} \nu &= \bar{\nu}[1+m\cos(nt-\phi)+m'\cos(2nt-\phi')], \\ \text{i.e.}\quad (\nu-\bar{\nu}) &= \Delta\nu = \bar{\nu}[m\cos(nt-\phi)+m'\cos(2nt-\phi')], \end{aligned}\right\} \tag{28}$$

where it has been assumed that $2m \ll 1$.

Equation (28) shows how ν oscillates about its mean value $\bar{\nu}$.

Suppose a second radio wave (wanted wave) to traverse a short path s over which $\bar{\nu}$ is effectively constant but with ν modulated as in (28).

Let ρ nepers be the absorption of this wave in the distance s so that if $(E_2)_0$ is the amplitude of the wave at the beginning of s then its amplitude at the end of s is:

$$\begin{aligned} E_2 &= (E_2)_0\exp(-\rho) = (E_2)_0\exp\left[-\bar{\rho}-\frac{d\rho}{d\nu}\Delta\nu\right] \\ &= (E_2)_0[\exp-\bar{\rho}]\left[1-\frac{d\rho}{d\nu}\Delta\nu\right] \\ &= (E_2)_0(\exp-\bar{\rho})\left[1-\left(\frac{d\rho}{d\nu}\right)\bar{\nu}\{m\cos(nt-\phi)+m'\cos(2nt-\phi')\}\right]. \end{aligned} \tag{29}$$

The second wave therefore becomes amplitude modulated at the frequencies of the fundamental and of the octave of the interfering wave. The amplitude of the transferred modulation at the fundamental and octave tones acquired along the path s are, respectively,

$$\left.\begin{aligned} M &= m\bar{\nu}\frac{d\rho}{d\nu} = M_0\bigg/\left[1+\left(\frac{n}{G\nu}\right)^2\right]^{\frac{1}{2}} \\ \text{and}\quad M' &= m'\bar{\nu}\frac{d\rho}{d\nu} = \frac{DM_0}{4}\bigg/\left[1+\left(\frac{2n}{G\bar{\nu}}\right)^2\right]^{\frac{1}{2}}, \end{aligned}\right\} \tag{30}$$

where, from (8),

$$M_0 = \left(\frac{d\rho}{d\nu}\right)\frac{w_0D}{\bar{Q}G}. \tag{31}$$

If $\bar{Q}G$ in (31) be replaced by its value in (24), then (31) becomes

$$M_0 = \left[\bar{\nu}\left(\frac{d\rho}{d\nu}\right)D\right] \Big/ \left[1 + D^2/2 + \frac{Q_t G\bar{\nu}}{w_0}\right]. \tag{32}$$

In practice $\frac{Q_t G\bar{\nu}}{w_0} \gg 1$, so that (32) is effectively

$$M_0 = \left(\frac{d\rho}{d\nu}\right)\frac{w_0 D}{Q_t G}.$$

When the frequency $p_2/2\pi$ of the wanted wave is such that $p_2^2 \gg \bar{\nu}^2$ then it is known that $\rho \propto \nu = \frac{CN\nu}{p_2^2}$ where C is a constant. It follows that $\bar{\nu}\frac{d\rho}{d\nu} = \bar{\rho}$, so that (31) becomes

$$\begin{aligned} M_0 &= \frac{\bar{\rho} w_0 D}{\bar{Q}(G\bar{\nu})} \\ &= \frac{CNw_0 D}{p_2^2 \bar{Q} G} \end{aligned} \tag{33}$$

in which N is the electron concentration.

The harmonic component M' in (30) has been observed (Huxley *et al.* 1947) to be present in the expected proportion.

Expression (31), and its limiting form (33), are given in a form which makes it possible to assess the effect of changes in any one of the contributory physical processes on the total transferred modulation M_0.

The phenomenon of the interaction of radio waves thus possesses a double interest, both as a link between laboratory and ionospheric studies and as a tool for ionospheric research. By use of this phenomenon it has been possible for the first time to make controlled alterations in the properties of the ionosphere itself.

REFERENCES

Best, N., Havens, R. & La Gow, H, 1947 *Phys. Rev.* **71**, 915.

Healey, R. H. & Reed, J. W. 1941 *The behaviour of slow electrons in gases.* Amalgamated Wireless (Australasia) Ltd.

Huxley, L. G. H. 1937 *Phil. Mag.* (7), **23**, 442.

Huxley, L. G. H. 1940 *Phil. Mag.* (7), **29**, 313.

Huxley, L. G. H., Foster, H. G. & Newton, C. C. 1947 *Nature*, **159**, 300.

Huxley, L. G. H., Foster, H. G. & Newton, C. C. 1948 *Proc. Phys. Soc.* **61**, 134.

Ratcliffe, J. A. & Shaw, I. J. 1948 *Proc. Roy. Soc.* A, **193**, 311.

Preparation and properties of a reproducible neutron standard

By E. Bretscher, G. B. Cook, G. R. Martin and D. H. Wilkinson

(*Communicated by Sir James Chadwick, F.R.S.—Received* 6 *August* 1948)

Radium beryllium fluoride ($RaBeF_4$) is proposed as a new form of radium-α-beryllium neutron source, for adoption as a standard. Three such sources have been prepared and calibrated; they give a neutron yield which is proportional to the radium content to within $\frac{1}{2}$%. Details are given of the techniques used in the preparation and standardization of these sources; probable errors in the calibration are discussed.

1. Introduction

In various nuclear physical experiments it is necessary to know the absolute yield of a neutron source. Several well-established procedures are available for this determination, but they are, without exception, lengthy, and require considerable preparation. If there were available neutron sources whose yields were accurately reproducible, then, after determination of the absolute yield in several laboratories, this problem would reduce to a comparison of the working source with such a standard. Such relative measurements are much simpler and quicker to make, and are susceptible of greater accuracy. In addition, the existence of such sources would considerably simplify the intercomparison of neutron measurements between various laboratories.

The characteristics desirable in such a source are:

(1) Ease of preparation.

(2) Independence from arbitrary parameters, such as distances, volumes, weight ratios, etc. (Ideally the source should be completely specified in terms of one arbitrary quantity, which would normally be the weight of active material used; all other variables should produce only second-order effects.)

(3) The neutron yield should be independent of time.

(4) Small physical dimensions (i.e. approximating to a point-source).

(5) The absolute value of the neutron yield must be reasonably high, to facilitate comparison with other sources.

The normal type of 'mixed' radium-beryllium source, although providing a high absolute yield, is known to be unreliable (Bright *et al.* 1941). The chief cause of variation appears to lie in the difficulty of achieving a homogeneous mixture of two powders (usually radium sulphate or bromide and beryllium metal) of such different densities, and of preventing partial separation under the influence of gravity. To avoid the last defect, Anderson & Feld (1947) have recently proposed compression of the mixture in a hydraulic press. This would eliminate at least the possibility of gravitational separation. Since the experiments described below were performed, Gamertsfelder & Goldhaber (1946) have proposed a 'reproducible neutron standard' which is based on the photo effect in Be, using Ra as a source of γ-rays. We had also considered this possibility, but came to the conclusion that such a source is rather sensitive to geometry, to corrections for γ-ray absorption in the

Ra-containers and in addition has a low efficiency (about 62 neutrons/sec./millicurie Ra). If reproducibility to within 1 % is aimed at, the necessity to correct for every possible error makes the method based on the photo effect less attractive than it might seem at first sight.

In order to overcome these objections, we have searched for a definite chemical compound between an α-active element (preferably radium) and a suitable target element such as beryllium or boron. An inspection of the literature disclosed no properly characterized radium compound containing either of these elements. The other three alkaline earth metals, Ca, Sr, Ba, however, yield salts of the type $Me^{II}BeF_4$ (where Me^{II} is a divalent cation), isomorphous with, and in many other respects similar to, the corresponding sulphates. It thus appeared a reasonable extrapolation to expect that radium beryllium fluoride would be a stable insoluble salt, similar to radium sulphate, which is the least soluble of the alkaline earth sulphates.

In a source of this type the following additional conditions have to be met:

(1) The dimensions of the source must be large compared to the range of α-particles in the material.

(2) The grain size must be large compared to the range of the recoil fragments resulting from α-emission (i.e. a low 'emanating power').

(3) The composition must be constant.

(4) The contribution of the polonium α-particles must be small, since this part of the neutron yield will grow with the half-life of radium D (22 years) with initially pure radium.

We believe that all of these conditions are substantially met in the sources we prepared; the magnitude of the uncertainties attributable to them is discussed later.

2. EXPERIMENTS WITH BARIUM BERYLLIUM FLUORIDE

(1) *Preparation and analysis*

In view of the difficulty and expense involved in handling radium salts, preliminary experiments were carried out on the corresponding barium compound, with the primary object of checking the constancy of composition of the complex, under varying conditions of preparation.

Stock, Praetorius & Priess (1925) have described the preparation of $BaBeF_4$, by two methods: (*a*) by precipitation from aqueous solutions of K_2BeF_4 and $BaCl_2$, and (*b*) by fusing together $BaCl_2$ and K_2BeF_4. Samples were prepared under various conditions, and their purity determined. In some cases complete analyses were carried out, but for most samples the ratio

$$\frac{BaSO_4 + BeSO_4}{BaBeF_4}$$

obtained by fuming with concentrated sulphuric acid to constant weight, was taken as a criterion of purity. The other analytical techniques employed were:

(*a*) For barium: the complex was decomposed with concentrated sulphuric acid in a platinum crucible, and the resulting mixture of sulphates extracted with dilute sulphuric acid. The remaining barium sulphate was dried and weighed.

(*b*) For beryllium: the complex was transformed into the sulphate mixture, and the beryllium determined iodometrically by the method of Bleyer & Moormann (1912).

(*c*) For fluorine: the complex was decomposed by heating with perchloric acid, water being added from time to time to keep the boiling-point at 110 to 115° C (Willard & Winter 1933). The distillate containing the hydrogen fluoride was concentrated and titrated with potassium aluminium sulphate solution using phenolphthalein + methyl orange as indicator (see Kurtenacker & Jurenka 1930).*

TABLE 1. RESULTS OF ANALYSIS OF $BaBeF_4$

preparation	Ba (%)	Be (%)	F (%)	sulphate ratio $\frac{BaSO_4+BeSO_4}{BaBeF_4}$
calculated for $BaBeF_4$	61·8	4·05	34·2	1·522
(1) $BaBeF_4$ prepared by fusion at 800° of $BaCl_2$ in an excess of K_2BeF_4, purified by washing with water several times	58·7 59·8	3·90 3·92	33·3 34·2	—
(2) $BaBeF_4$ prepared by precipitation from neutral cold aqueous solution with 1 % K_2BeF_4 solution	61·3 62·0 61·1	4·00 4·07 4·09	35·3 34·4 34·1	1·524
(3) $BaBeF_4$ prepared as (2), but after precipitation the solution was brought to the boil and cooled	—	4·02 4·10	—	1·523 1·515
(4) $BaBeF_4$ prepared as (2), but the solution boiled for 15 min.	—	4·09	—	1·529 1·516
(5) $BaBeF_4$ precipitated from boiling N/20-HCl solution using 1 % $BaCl_2$ and K_2BeF_4 solutions	—	—	—	1·519
(6) As (5), but in N/5-HCl, the K_2BeF_4 solution added quickly	61·5	—	—	1·522 1·519 1·522
(7) As (6), but the K_2BeF_4 solution added slowly drop by drop	—	—	—	1·521

These analytical procedures were first checked on standard solutions of the elements concerned and found to be satisfactory. The results obtained with samples prepared in various ways are listed in table 1. It is clear from this that the fusion method is not satisfactory, and such a method would in any case be objectionable when the radium compound had to be prepared. Within the experimental error, the samples prepared from aqueous solutions all appeared to possess the required composition.

To find out if any of the Ba was precipitated as BaF_2, X-ray powder patterns of the samples of $BaBeF_4$ were compared with patterns obtained from samples of our purest preparation adulterated with 1, 3 and 10 % of barium fluoride (the most likely impurity). The method proved to be insufficiently sensitive, the limit of detection being about 3 %, and the results were thus not very significant. It can be said, however, that our samples obtained from aqueous solutions, contained less than 3 % of BaF_2. The sample obtained by fusion appeared to contain about 10 % in rough agreement with the analytical figures already quoted.*

---

* We are indebted to Dr C. J. Wilkins for help with the determination of fluorine.

(2) *Co-precipitation experiments*

Evidence for the existence and nature of the corresponding radium salt was obtained by radioactive tracer methods. A study of the distribution of traces of mesothorium (Ra^{228}) between precipitated barium beryllium fluoride and the solution after partial precipitation by insufficient amounts of K_2BeF_4, showed that the radium salt is, in all probability, less soluble than the barium compound. This was taken as confirmation of the initial assumptions about the stability and solubility of the radium compound.

(3) *Emanating power of the complex*

It is a condition for the reproducibility of the source that the emanating power of the preparation should be low. A small quantity of radium chloride was mixed with barium chloride solution and samples of barium beryllium fluoride were prepared in various ways. These samples were allowed to emanate in an atmosphere of hydrogen, which was later pumped off through a charcoal trap immersed in liquid air, to absorb the radon. The γ-activity of the charcoal tubes was later measured and compared with that from an aliquot of the original solution. Table 2 gives the percentage emanating power of the samples. These figures are in qualitative agreement with the conclusions reached from a comparison of the crystal sizes made under a low-power microscope.

TABLE 2. EMANATING POWER OF $RaBeF_4$ COPRECIPITATED WITH $BaBeF_4$

conditions of precipitation	emanating power (%)
cold, neutral, 1 % solutions	1·5
	1·8
boiling, neutral, 1 % solutions	2·7
	2·0
boiling, 1 % solutions in N/10-HCl	0·3
	0·1
boiling, 1 % solutions in N/5-HCl	0·1
	<0·1

3. PREPARATION OF RADIUM BERYLLIUM FLUORIDE†

About 40 mg. of radium, as bromide, were available for the preparative work. It had recently been specially purified, and the purity was given as > 99 %. The salt was weighed into three roughly equal portions, and an independent preparation made with each in the following manner:

The salt was dissolved in 1 ml. of N/5-HCl in a small platinum crucible, and heated on a hot plate. To the nearly boiling solution was added 2·5 ml. of boiling K_2BeF_4

---

* We are indebted to Dr A. J. C. Wilson, then of the Crystallographic Laboratory, Cambridge, for making these measurements.

† We are very much indebted to Messrs Thorium Ltd., who supplied the radium bromide used, for the provision of laboratory facilities for the work with radium. In addition, we have to thank Dr W. P. Grove for advice on radium handling technique, and for making the emanation measurement on the final material.

solution (containing 15 mg. K_2BeF_4, about 2½ times the theoretical quantity) dissolved also in N/5-HCl. A white precipitate formed and quickly settled. After cooling, the supernatant liquid was removed with a dropper and teat, without disturbing the precipitate, which adhered strongly to the platinum crucible, and centrifuging was therefore unnecessary. The precipitate was washed three times with hot water, the washings being removed as before, and the solid then dried for 1 hr. at 140° C. The $RaBeF_4$ was removed from the walls of the crucible by means of a platinum microspatula coated with 'Polythene' to prevent scratching of the crucible.

Samples were removed from the first two preparations and analyzed by fuming to constant weight with concentrated sulphuric acid.

		I	II
ratio $\frac{RaSO_4+BeSO_4}{RaBeF_4}$	found	1·378	1·380
	calc.	1·373	

The third sample was used first for the measurement of the emanating power and gave a result of about 3 %; the significance of this will be discussed later. This sample, and the remainders of the first two were then weighed into small platinum capsules and sealed. The capsules were identical cylinders, 0·37 cm. high, 0·25 cm. in diameter, and with wall thickness 0·025 cm. The radium content was determined by comparison of the γ-ray ionization currents with that of a radium standard:

	I	II	III
wt. $RaBeF_4$ (mg.)	11·162	8·543	—
wt. Ra(calc.) (mg.)	8·11	6·21	—
wt. Ra (from γ-ray measurements)*	8·03	6·13	18·04

* A correction for absorption of the γ rays by the platinum container has been applied.

The discrepancy between the calculated and measured weights of radium is of the same order as the experimental error, but may be explained by the presence of ~1 % Ba in the radium used.

4. Inter-comparison of the three sources†

Comparisons were made between the three sources by activation of a manganese detector with the neutrons from each one in turn, after slowing down in a large water bath. The detector was powdered electrolytic manganese, packed into the annular space between two thin brass cylinders. The counter used was of the thin-walled, self-quenching type, with a sensitive volume of length 7·0 cm., diameter 1·80 cm.; the walls were of copper 0·012 cm. thick. The counter sensitivity was checked with a source of UX_2 in equilibrium with uranium, and the correction which had to be applied to the observed results never exceeded 0·25 %. By adopting a standard counting routine any error arising from the uncertainty in the Mn^{56} half-life was eliminated. Under these conditions, it was assumed that the Mn^{56} activity, after correction for change in counter sensitivity, for counter background, and for finite time of irradiation, was proportional to the total neutron flux from the source.

---

† A more detailed account of this and the succeeding section will be published separately.

These measurements were carried on until sufficient statistical accuracy had been achieved (at least 10^5 disintegrations were recorded in each case), and gave the following results:

source no.	I	II	III
Ra content (mg.)	8·03	6·13	18·04
relative γ-ray activity	1	0·765 ± 0·002	2·247 ± 0·005
relative neutron emission	1	0·767 ± 0·003	2·246 ± 0·004

These results show that the neutron emission is proportional to the amount of radium to within 0·5 %. They thus establish the suitability of radium beryllium fluoride for the preparation of standard neutron sources.

5. ABSOLUTE NEUTRON YIELD OF $RaBeF_4$ SOURCE

An approximate absolute value for the neutron yield of our sources was obtained by comparison with two previously standardized 'mixed' sources in the same arrangement. Before doing this, two important points were checked:

(*a*) The 'mixed' source was considerably larger in physical dimensions than our $RaBeF_4$ sources, and raised the possibility that the efficiency of the detecting system would vary from point to point within it. Experiments were therefore performed in which the detector was activated by the small source placed at various positions within the cylinder, and a 'plateau volume' was mapped in which detector efficiency was constant. This volume was big enough to accommodate the larger source.

(*b*) The neutron spectrum of the two types of source would be different owing to the neutron contribution from fluorine in the one case only. The thermal neutron density in the slowing-down medium was therefore investigated for the two types of source, with a small BF_3 filled chamber. Close agreement was found between the two types, and the error introduced by assuming that the slow neutron density at the detector is proportional to the fast neutron emission from the source, is thought to be negligible.

The previous standardization of the 'mixed' sources had been carried out by irradiation of a large tank of manganous sulphate solution (Amaldi & Fermi 1936), first with this source, and then with the D + D source of the Cambridge high voltage equipment. The neutron flux from the D + D source was measured with an ethylene-filled, 'Polythene'-lined ionization chamber and valve electrometer. It is possible that the accumulated errors in this chain of experiments may amount to as much as 10 %.

The value obtained was $1{\cdot}84 \times 10^6$ neutrons/sec./g. of $RaBeF_4$; this was the weighted mean of comparisons with two 'mixed' sources. The magnitude of this yield is most gratifying in view of the small Be : Ra ratio which these sources show. The ratio normally utilized in 'mixed' sources is of the order of 200 : 1 (by atoms). Such a source is quite small and compares favourably with the photoneutron source referred to by O'Neal & Scharff-Goldhaber (1946), which gives $8{\cdot}6 \times 10^5$ neutrons/g. radium surrounded by 85 g. Be.

6. DISCUSSION OF ERRORS

A number of factors may contribute to a lack of reproducibility between different preparations of the source. The sources of error to be considered are:

(i) *Loss of radon from the crystals*

In one source approximately 3 % of the radon escaped from the solid phase. This will not, of course, escape from the platinum capsule, but its α-particles may not be so efficient in neutron production as those originating from radon retained within the crystals. This effect would not be very large, and will be further mitigated by the compactness of the source. Moreover, the value of 3 % emanating power was obtained when the $RaBeF_4$ was spread thinly over the walls of the platinum crucible, so that the emanating power might have been lower if the powder had been packed into a small container.

(ii) *Loss of α-particles into the platinum walls of the container*

The surface layer of the source will emit some α-particles whose path lies partly in the platinum walls, and whose efficiency is therefore less than that of the rest of the α-particles. The relative magnitude of this loss will clearly depend on the total weight of source material. A rough calculation has been made, assuming a spherical source, and taking the excitation function for the Be(α, n) reaction obtained by Bjerge (1938). This should give at least the order of magnitude of the deviation of the neutron yield from strict proportionality to the source weight.

TABLE 3. NEUTRON LOSSES DUE TO WALL EFFECT BY RADIUM AND DECAY PRODUCTS

I	II	III	IV	V	VI	VII
Ra	3·25	2·23	0·016	0·052	0·244	0·02
Rn	4·05	2·78	0·020	0·111	0·238	0·05
Ra A	4·65	3·20	0·023	0·181	0·232	0·10
Ra C′	6·90	4·75	0·034	0·565	0·258	0·50
Po	3·83	2·63	0·019	0·091	0·234	0·04
						0·71 %

I α emitter involved.
II range of α's in cm. of standard air.
III range of α's in $RaBeF_4$ (cm. × 10^{-3}), assuming a mean atomic stopping power of 1·6.
IV fraction of α-particles whose paths end outside the source material (calculated for a 20 mg. sphere of $RaBeF_4$).
V relative neutron-producing efficiency of α-particles of range given in (III) (from data of Bjerge 1938).
VI average neutron-producing efficiency of a family of α-particles with rectangular distribution of ranges (maximum = III).
VII weighted percentage loss of neutrons (= (IV) × (V) × (VI) × 100).

It will be obvious that α-particles can escape only from a superficial spherical shell of thickness equal to their range and calculation shows that of all the α-particles emitted from atoms in such a shell, three quarters will, in any case, end their paths within the solid, and that the remaining quarter will have a rectangular distribution of residual ranges (if the radius of the sphere is $\gg$ the range of the α-particle).

Bjerge's excitation function was obtained with a thin target, and a thick target curve was obtained by graphical integration; from this the quantities given in table 3 were evaluated.

Table 4 gives the magnitude of the correction to be applied to each of our three sources. It will be seen that none of them is subject to a correction on this account as large as 1 %, and since the error diminishes with increasing mass, it cannot be regarded as serious.

TABLE 4. CALCULATED CORRECTION TO $RaBeF_4$ SOURCES FOR WALL EFFECT

source	weight (mg.)	correction (%)
I	11·162	0·85
II	8·543	0·94
III	25·00	0·66

(iii) *Growth of polonium in the source*

Polonium will grow in initially pure radium with, roughly, the half-life characteristic of RaD (22 years). Table 3 shows that about 9·1 % of the neutron yield of a source in equilibrium will be due to Po α-particles, and will grow, therefore, in this way. The remainder of the neutron yield will effectively attain its maximum value within one month. A correction must therefore be made for the growth of polonium, amounting initially to 0·28 % per annum. The values which we quote for the neutron yield of the source were obtained before any significant growth of Po had taken place.

(iv) *Chemical composition*

Our experiments with the barium compound (see table 1), and the two analyses of the radium compound, have been taken as sufficient proof of its stoichiometric nature.

This work was carried out during 1943 on behalf of the Directorate of Tube Alloys of the Department of Scientific and Industrial Research. The paper has been published with the permission of the Director of the Atomic Energy Research Establishment, Ministry of Supply. This is a concentrated version of the original report B.R. 383.

REFERENCES

Amaldi, E. & Fermi, E. 1936 *Phys. Rev.* **50**, 899.
Anderson, H. L. & Feld, B. T. 1947 *Rev. Sci. Instrum.* **18**, 186.
Bjerge, T. 1938 *Proc. Roy. Soc.* A, **164**, 243.
Bleyer, B. & Moormann, A. 1912 *Z. anal. Chem.* **51**, 360.
Bright, W. C., Clarke, E. T., Korff, S. A., Murphy, E. J. & Whitacker, M. D. 1941 *J. Franklin Inst.* **231**, 357.
Gamertsfelder, G. R. & Goldhaber, M. 1946 *Phys. Rev.* **69**, 368.
Kurtenacker, A. & Jurenka, W. 1930 *Z. anal. Chem.* **82**, 210.
O'Neal, R. D. & Scharff-Goldhaber, G. 1946 *Phys. Rev.* **69**, 368.
Stock, A., Praetorius, P. & Priess, O. 1925 *Ber. dtsch. chem. Ges.* **58**, 1571.
Willard, H. H. & Winter, O. B. 1933 *Industr. Engng Chem.* (Anal. ed.), **5**, 7.

The kinetics of the interaction of atomic hydrogen with olefines

I. Apparatus and use of para-hydrogen techniques

By H. W. Melville, F.R.S. and J. C. Robb

Chemistry Department, Marischal College, Aberdeen

(*Received* 23 *June* 1948—*Read* 10 *February* 1949)

It is shown in this paper that the normal techniques and methods of approach to an investigation of radical reactions are not applicable in dealing with reactions involving the addition of a hydrogen atom to an olefine, since this type of reaction is so fast that even the exchange conversion of para-hydrogen by hydrogen atoms is inhibited by very small amounts of olefines.

A new type of reaction has been utilized to provide a means of competing with the olefine for a hydrogen atom, namely, the removal of a hydrogen atom on a layer of molybdenum oxide or tungsten oxide. In order to adapt the system to the measurement of reactions over a wide range of collision efficiency, a special reaction vessel is described in which the path length offered to a hydrogen atom before removal on the oxide layer can be varied.

A colorimetric technique has been devised for measuring the rate of addition of hydrogen atoms to the oxide surface. This technique is shown to be very sensitive, and the addition of a very small amount of hydrogen atoms can be detected.

The para-hydrogen conversion technique has been applied to illustrate how the collision efficiency of a hydrogen atom with a saturated hydrocarbon molecule can be obtained. The reaction $H+RH=H_2+R$ has approximately the same collision efficiency as the reaction $H+p\text{-}H_2=n\text{-}H_2+H$, and hence in the presence of a saturated hydrocarbon there is no inhibition of the para-conversion, only retardation.

A further application of the para-conversion has been to demonstrate the efficiency of the removal of a hydrogen atom on the oxides of molybdenum and tungsten.

INTRODUCTION

The processes involved in the thermal and photochemical reactions of gaseous hydrocarbons are now well recognized to consist of radical reactions of various types. If, therefore, these processes are to be properly understood, it is necessary to build up a chemistry of the appropriate radical reactions with especial emphasis on the quantitative aspect. This latter feature is perhaps more important than has been realized, because the overall character of a reaction sequence is determined by the absolute magnitude of the coefficients of the possible intermediate radical processes.

There is now quite a variety of methods of generating radicals at a well-defined rate or at a rate which can be measured by suitable techniques. On the other hand, there is practically no definite knowledge about the ways in which the radicals disappear. For example, although it is known that methyl and higher alkyl radicals can combine on the walls of a reaction vessel, presumably by processes akin to those which operate in the case of free atoms, there is no knowledge regarding gas-phase removal of radicals. Since the rate of removal of radicals is unknown, it is consequently impossible to calculate their stationary concentration even under the simplest of conditions.

In spite of this severe limitation, however, a number of chemical reactions of free radicals have been investigated, both energies of activation and temperature independent factors being obtained. Even if absolute values are difficult to get, it is often possible to calculate relative values with a satisfactory degree of accuracy. The intensive progress and also the wide discrepancies between the results of various authors are well brought out in E. W. R. Steacie's book on *Free radical chemistry in gases*. Here, again, there is a very awkward situation. Suppose a radical R_1 reacts with a hydrocarbon $R_2\mathrm{H}$ as follows:

$$R_1 + R_2\mathrm{H} = R_1\mathrm{H} + R_2,$$

then, in estimating rates, it is necessary to measure chemically the rate of accumulation of $R_1\mathrm{H}$. Unfortunately, the radical R_2 may undergo a series of reactions which may so considerably complicate the chemistry of the whole process that it makes it very difficult indeed to investigate one individual process to the exclusion of all others. What is required is some method of removing the radical R_2 from the sphere of action.

From general experience it is known that atom-saturated molecule or radical-saturated molecule reactions involve rather high energies of activation, and therefore there is no particular difficulty in measuring these velocities by existing well-explored techniques such as the exchange conversion of para-hydrogen

$$\mathrm{H} + p\text{-}\mathrm{H}_2 = o\text{-}\mathrm{H}_2 + \mathrm{H},$$

to which further reference will be made. With unsaturated molecules, however, matters are otherwise, the reactivities being such that the energies of activation are very considerably lowered—so much so that they are not easily accessible to direct measurement.

The present series of papers is designed to give a description of an approach to an investigation of the collision efficiency of a hydrogen atom with an olefine by a new technique which has yielded results which indicate the potentialities of this method in the investigation of a wide range of the elementary reactions of hydrocarbons.

APPARATUS

The apparatus was constructed entirely of soda glass. A general view of the arrangement of the main vacuum system, the reaction system and the analytical systems is seen in figure 1.

The pumping equipment consisted of an Edwards type 6 double-stage mercury diffusion pump, backed by a single-stage Speedivac oil pump.

Along the whole length of the grid ran the main vacuum line, on which was a barometer-type manometer to measure high pressures and a vacuostat reading from 10^{-3} to 1 mm. A subsidiary intermediate vacuum line was connected in the forevac system and also ran the whole length of the apparatus. This line was used for pumping out liquid-air containers when a temperature of less than 90° K was required. Since the rotary oil pump had to provide a backing pressure for the mercury pump and also to be used for pumping out liquid-air vessels, it was desirable to introduce a 6 l. bulb in the forevac system so that the mercury pump could operate without the necessity for having the oil pump in continual operation.

The rest of the apparatus can be conveniently discussed under the following three headings:

(*a*) Gas storage and purification,
(*b*) Reaction system,
(*c*) Analytical systems.

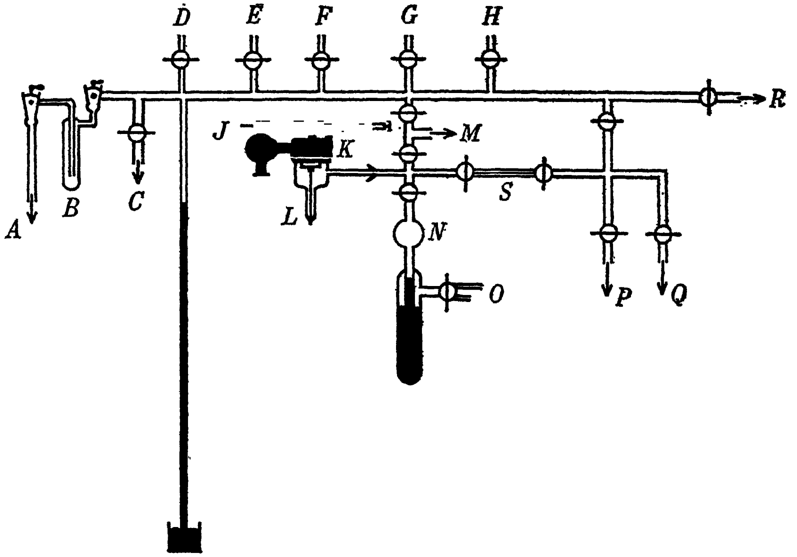


FIGURE 1. General arrangement of apparatus. *A*, to vacuum pumps; *B*, liquid-air trap; *C*, to vacuostat; *D–H*, to gas reservoirs; *J*, mercury lamp; *K*, photoelectric colorimeter; *L*, short-path reaction vessel; *M*, to vacuostat; *N*, injector; *O*, to water pump; *P*, to olefine analyzer; *Q*, to thermal conductivity gauge; *R*, to methane and olefine production units; *S*, capillary dose lock.

Gas storage and purification

The gases were stored in 2 l. bulbs, all of which were fitted with barometer-type manometers. The gases stored were: (*a*) hydrogen and *p*-hydrogen, (*b*) ethylene, propylene, butylene, methane and other higher boiling hydrocarbons.

Since the first group of gases had to be used for the calibration of the apparatus, small doser taps of the three-way variety were fitted to these reservoirs. The other gases had provision for condensing to remove condensible fractions and to permit outgassing of the liquid hydrocarbons used.

Gas purification and preparation

Cylinder hydrogen was purified by diffusion through a palladium thimble at a temperature of about 500° C.

Para-hydrogen was prepared by the usual low-temperature method. Several variations of this method were tried out. One method was to absorb the hydrogen on shell charcoal at 65° K, maintain contact for about an hour, and then allow the hydrogen to desorb on warming the charcoal. The content could not, however, be raised above 30 %. Increasing the time of absorption on the charcoal at low temperatures did not significantly increase the yield. A flow method under the same conditions was tried, but still no increase in the para content was observed.

A new variation of the method, however, proved much more satisfactory. It consisted of impregnating a sample of shell charcoal with dysprosium oxide, which it was hoped would prove to be a satisfactory flow catalyst in view of the large magnetic moment of dysprosium.

The catalyst was prepared by dissolving 0·86 g. dysprosium oxide in dilute nitric acid and evaporating down the solution on a water-bath. Sufficient cold water just to dissolve the nitrate was added, and 10 g. of shell charcoal (B 570 V.A. 17·0 $\frac{7}{18}$ mesh) was introduced and more water added just to moisten the charcoal. The water was then evaporated off on a water-bath, dried in an air-oven at 130° C for $2\frac{1}{2}$ hr. and calcined in a limited supply of air at 350° C in an electric furnace. The catalyst was then outgassed *in vacuo* at 350° C. By flowing hydrogen through this catalyst (at 65° K) at the relatively fast rate of 30 ml./min., the para content could be raised to 36 %. This was the most satisfactory method of all.

Methane was prepared for calibration purposes only, since it was considered that one of the products of reaction of hydrogen atoms with olefines or with the hydrocarbons formed by secondary reactions might be methane. In the early stages of the work, it was feared that any methane which might be formed would interfere seriously with the measurements of the thermal conductivities of the hydrogen samples.

The methane was prepared by the action of aluminium carbide on hot water. The gas was then scrubbed by passing it through copper and copper oxide at 360° C, strong caustic soda solution, dried by passing through activated alumina, and purified by condensing the high-boiling constituents in a solid CO_2-acetone bath. The product of this treatment was then condensed in liquid air and a suitable distilled middle fraction retained.

Ethylene of the commercial variety was further purified by outgassing thoroughly and distilling; the middle fraction only being retained.

Propylene and iso-*butene* were prepared by the catalytic dehydration of *n*-propyl alcohol and *tert*.-butyl alcohol on activated alumina at 340 to 370° C. The products were dried and distilled from liquid air, again the middle fractions only being retained.

Molybdenum oxide was prepared by the method suggested by Liebert (1931), commencing from ammonium molybdate.

A 50 g. portion of ammonium molybdate was boiled with 100 ml. aqua regia on a steam bath until no ammonia could be detected by Nessler's reagent. The resulting solution of molybdic acid was then evaporated down almost to dryness and then twice more with fresh acid. The solid residue was dried in an air oven at 180° C. The oxide as prepared above was a light yellow almost white powder, which on exposure to hydrogen atoms turns blue. Three methods were adopted in regenerating the white form from the blue: (*a*) evaporation twice with dilute nitric acid followed by drying as before, (*b*) by evaporating once with 0·880 ammonia, then with aqua regia as before, (*c*) by heating to about 360° C in an electric furnace for several hours.

The last method is of all the easiest to carry out in practice and also gives a cleaner product. A sufficiency of oxygen is required for this latter treatment, and a slow current of air circulated over the powder while being heated makes the whole process much faster.

Tungstic oxide was prepared from tungstic acid by heating to 350° C until all water was removed. The oxide was obtained as a rather gritty powder, yellow in colour.

Reaction systems

In this investigation, two entirely different types of reaction vessel were used. The first was a conventional type of reaction vessel, cylindrical in shape, diameter 6 cm. and depth 2 cm. This vessel was entirely of silica.

The other vessel was designed to meet the special needs of this problem and merits further discussion and explanation.

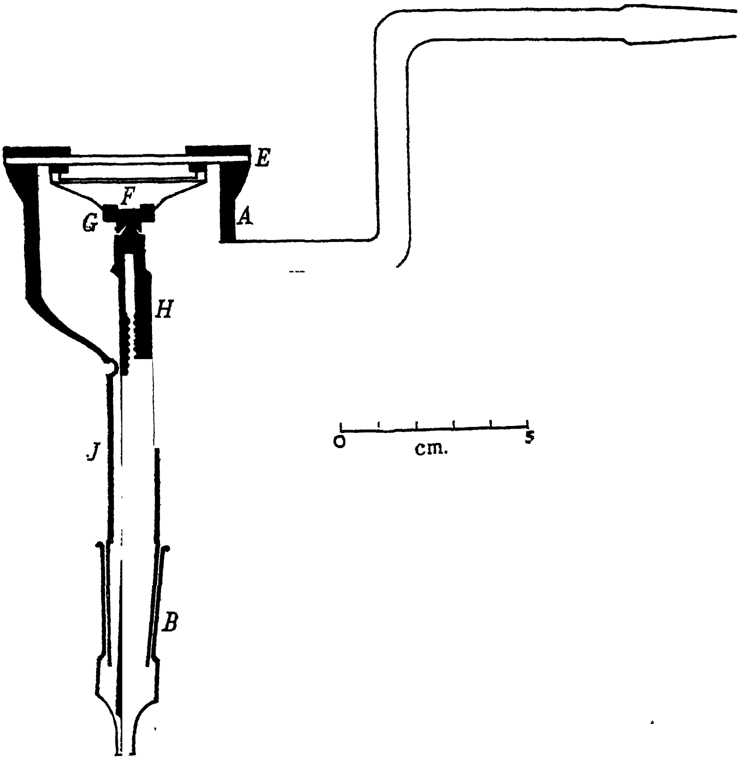


FIGURE 2. Short-path reaction vessel.

A diagram of this vessel is shown in figure 2. The main body of the reaction vessel, *A*, was formed from the neck of a 2000 ml. Hysil bolthead flask which was chosen for the broad flange on top. The side tube was fitted with an A 10 ground-glass cone to facilitate removal of the whole reaction system from the rest of the apparatus for cleaning and adjusting, an operation which had to be carried out after each run. *F* is a shallow glass dish supported on four springs screwed into a brass holder *G*. The glass dish could be maintained at a fixed distance from the silica window *E* by placing three small blocks of glass of predetermined thickness on the rim of the dish. A ring of soft glass fitted on top of the silica window limited the entrance of radiation of wave-length 2537 Å to the volume directly above the dish *F*. When the apparatus was in operation, the dish *F* contained a layer of molybdenum oxide packed and smoothed to a surface level with the top edge of the dish.

The lower end of the brass spring holder *G* had a conical recess and fitted over a conical point on the brass column *H*. The purpose of this point support for the plate was that even if the column was not truly perpendicular to the silica window of the reaction vessel, the plate would still automatically line itself up parallel to the silica window at the predetermined spacing and would be securely held there by the tension in the four springs. The column *H* consisted of a brass tube screwed internally and with three milled flutes on the outside, the purpose of these flutes being to locate the tube and permit it to slide in the centre of the glass limb *J*, while at the same time allowing adequate space for diffusion from the lower part of the limb to the body of the reaction vessel. A small indentation in the glass limb *J* engaged in the space between two flutes and prevented the whole assembly from turning. A brass rod screwed externally to fit into the tube *H* had two flats filed on its lower end so that it fitted a similarly shaped hollow on the end of the ground glass cone *B*. By turning the cone *B* the whole internal part could be raised or lowered, and a path of variable

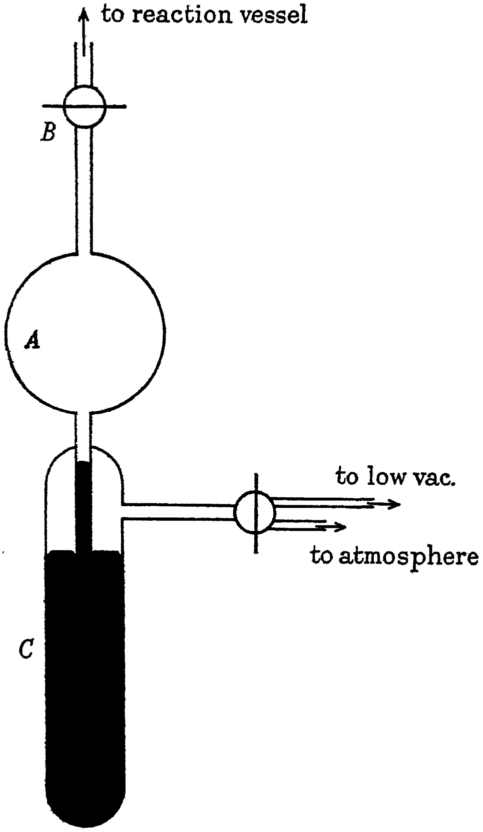


Figure 3. Gas injector.

distance could be provided at will between the silica surface and the molybdenum oxide contained in the shallow dish.

After some considerable use, the brass parts were found to be showing signs of attack by mercury. All brass fittings were then replaced by corresponding pieces made of duraluminium, which seemed to be more resistant to the attack of mercury vapour.

Gas injector

A type of injector, was constructed to assist in the accurate introduction of amounts of olefines to the reaction vessel in presence of hydrogen. It is shown in figure 3, and consisted of a bulb *A* connected to the reaction vessel system by means of a tap *B*. The bulb could be filled with mercury from the reservoir *C* by the usual technique used for mercury cut-offs. The volume relationship between the injector and the reaction system then gives the pressure of the gas which will be introduced to the reaction vessel when expelled completely from the injector at known pressure.

The operation consisted of introducing to the injector the required pressure of olefine, closing the tap *B* and evacuating the remainder of the reaction system. Hydrogen at known pressure could be introduced to the reaction vessel, and at the appropriate time, by opening tap *B* and running the mercury up until just past *B*, an accurately known pressure of olefine could be introduced. This piece of apparatus fulfilled a further function, since by running the mercury up and down several times before finally closing tap *B* the gases could be very thoroughly and efficiently mixed. There was then no need to wait for diffusion to establish a homogeneous mixture.

Analytical systems

Photoelectric colorimeter

As will be described later the basis of the method adopted in the present investigation was to estimate the rate at which hydrogen atoms added on to a molybdenum oxide surface. Since the surface turned blue on the addition of hydrogen atoms, the rate of blueing is a function of the rate of addition of hydrogen atoms. The apparatus for measuring the blueness of the surface is shown in figure 4. It consisted of a 25 W clear lamp, silvered lightly on the outside entirely with the

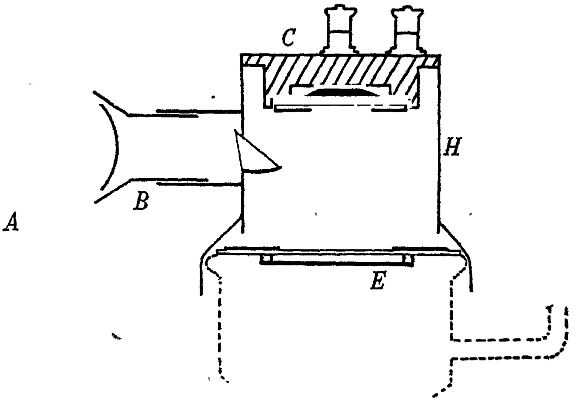


FIGURE 4. Photoelectric colorimeter.

exception of a small aperture about ½ in. in diameter, through which light projected along the tube *B*, which fits closely to the lamp and is made of two tubes fitting one within the other. The conical end can be made a light-tight fit on the lamp. At the end of this tube a small polystyrene prism-lens combination is let into the large brass tube *H*. The prism directs the light on to the molybdenum oxide surface in the special reaction vessel already described at *E*. The intensity of the reflected radiation was measured by means of a photocell *C*, and the whole assembly was rigidly mounted on a bar which was pivoted so that the whole instrument could be swung into position to make a determination.

When placed in position for use, the colorimeter was swung up against a metal rod and held firmly there by means of an elastic band. This ensured that the apparatus would occupy each time as nearly as possible the same position in relation to the reaction vessel.

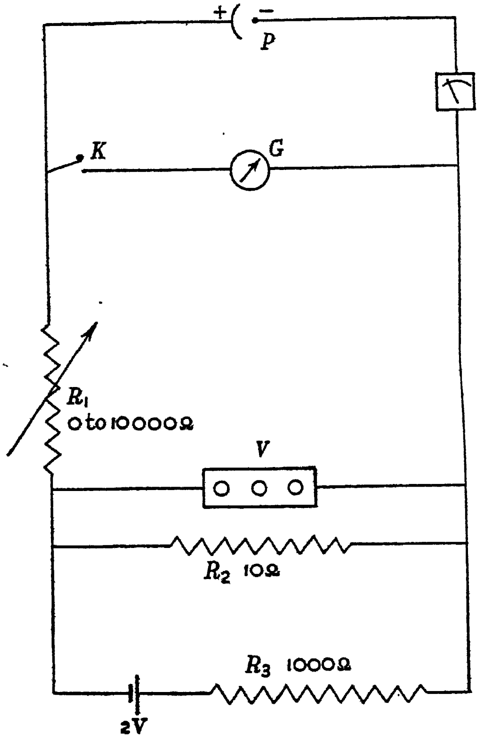


Figure 5. Circuit for measuring photocell current.

The photocell current was originally measured by a 0–60 microammeter of the unipivot type, but it was found that the mechanical limits set on the reproducibility of this instrument did not allow of precise enough measurements of the photocell current. It seemed that precision measurement of current by a single meter in the region of 10 μA was not possible, and some means had to be obtained of measuring the current while maintaining a low external resistance on the photocell. This low external resistance was an essential point, since high resistance introduced a marked fatigue effect from the cell, due to accumulation of electrons on one surface of the cell, and the inability of these electrons to leak through the high resistance fast enough to prevent this layer from forming.

In figure 5 is shown the circuit finally adopted for measuring the photocell current. The photocell P was connected via a low-resistance microammeter across a low-resistance galvanometer G. The microammeter is not an integral part of the circuit but was introduced so that a quick rough check of the photocell current could be obtained when necessary. When the galvanometer key K is pressed, the current flowing through the galvanometer is opposed by another current flowing through in the opposite direction, and the magnitude of this current is adjusted until a null point is reached. The balancing current is produced by using a potentiometer circuit on a 2 V accumulator. From this, a voltage of approximately 2×10^{-2} V is obtained across resistance R_2. This voltage is measured by a potentiometer voltmeter V and is connected across the galvanometer and a series resistance R_1, variable from 0 to 10,000Ω. The galvanometer resistance is about 10Ω, and so a variation from 10 to $10^4\Omega$ can be obtained. This gives a current range between 1 μA and 1 mA, the awkward range where pointer instruments are too insensitive and mirror instruments too sensitive. Thus by measuring the value of the voltage applied across R_1 and the value of R_1, the photocell current can be obtained precisely. The precision of this method is quite high, since a constant check is kept on the voltage dropped across R_2. The current can be measured easily and reproducibly to ± 0·15 % at about 13 μA.

The intensity of the 25 W lamp was maintained constant by running it off a voltage stabilizer of the electronic type.

Measurement of double-bond content in a sample of olefine and hydrogen

A method was developed for measuring the number of olefine molecules in a given sample. The apparatus is shown in figure 6. A small silica tube was fitted to the compression capillary of a McLeod-type gauge by means of a cone and socket. The sample which was to be analyzed was then compressed into the silica tube to a fixed mark and the pressure on the sample measured. A small low-pressure silica mercury-vapour lamp was then used to irradiate the sample, and the double bonds in the sample were hydrogenated by the photochemically produced hydrogen atoms. The volume of the sample was then adjusted to the mark as before and the new pressure recorded. The diminution of pressure was then a direct measure of the percentage double-bond content in the gas sample. An estimation could be completed in 5 min. with an error of ± 1 % of the olefine content in a sample containing about 20 % olefine.

Analysis of para-hydrogen mixtures

The gauge constructed for the measurement of the thermal conductivity of the hydrogen samples was of the type described by Bolland & Melville (1936) and later modified by Birse & Melville (1940*a*). It was, however, modified slightly, and a diagram of the gauge is shown in figure 7. The gauge was smaller than the types previously described, and a volume of 1 mm.^{3} at N.T.P. only was required to determine the thermal conductivity and para content. For ease of construction, the gauge was made with a horizontal filament instead of the more conventional vertical filament. The dimensions of the head of the gauge which contains the filament is about 15 mm. long and 1·5 mm. bore. The filament consisted of an unused coiled tungsten electric-lamp filament, 10 mm. long and 0·0745 mm. diameter, spot-welded to two pieces of borated copper wire. It is important that slight tension is put on the filament while sealing into the gauge, since repeated cooling and heating do not then tend to alter the geometry of the system. In this respect, the coiled filament is excellently suited for the purpose, and a further advantage lies in the fact that the gauge is also less pressure dependent than is the case when a straight wire is used.

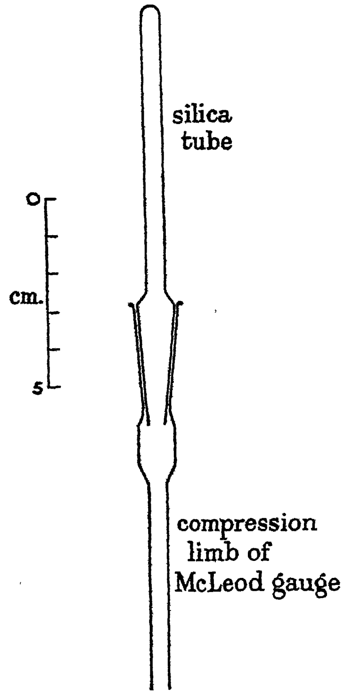


Figure 6. Apparatus for estimation of olefine content.

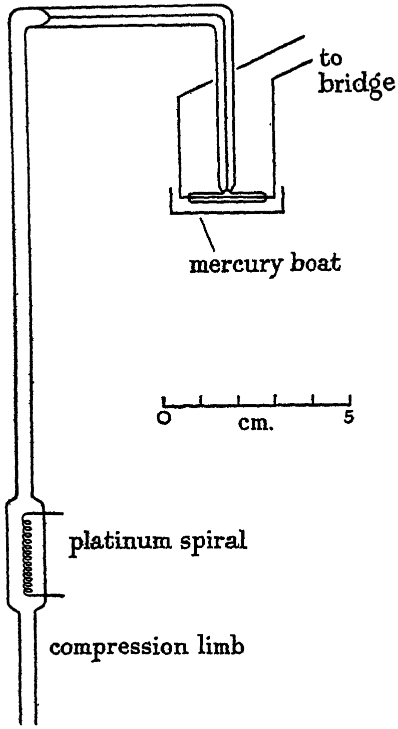


Figure 7. Micro-thermal conductivity gauge.

The gauge was operated immersed in liquid air by the usual technique of constant voltage, variable resistance measurement. When the gauge was first tried out, the balancing galvanometer in the Wheatstone net showed rapid and large fluctuations which completely upset the precision of the apparatus. It was discovered that the fluctuations were due to the local heating of the liquid oxygen on the wall of the gauge. The liquid oxygen then vaporized and collected as gas on the surface of the gauge. The gas collected into bubbles and finally broke away from the surface. This cycle was repeated, and the thermal discontinuity set up by the gas film on the gauge was responsible for the fluctuations of the temperature of the filament. This effect was entirely eliminated by immersing the gauge in mercury in a small steel boat as illustrated in the diagram. The leads to the gauge were insulated from the mercury by coating them with polystyrene. It was also found that provided liquid air was not allowed to cover the mercury block until the mercury had solidified completely, the mercury contraction did not on any occasion crush the glass.

The gauge was mounted on top of the compression capillary of a McLeod-type gauge and the pressure in the gauge adjusted to 50 mm. ± 1 mm. by means of a sliding millimetre scale fitted between the compression capillary and the comparison capillary.

To extend the scope of the gauge to the analysis of ternary mixtures, a platinum spiral was fitted to the compression capillary. This spiral could be glowed by connecting up to a 2 V accumulator. At a pressure of 0·4 mm., about 1 min. exposure to the glowing spiral was sufficient to convert the para-hydrogen to normal hydrogen. The frequent immersion of the equilibrating spiral in mercury did not seem to affect its catalytic efficiency. To keep the spiral really active, it was occasionally glowed for about an hour in an oxygen-hydrogen mixture, followed by a further 15 min. glowing in pure hydrogen.

Para-hydrogen technique

Although the para-hydrogen conversion technique was used only for the determination of the quantum input, there are some matters which require comment. The apparatus has already been described for the measurement of the thermal conductivity of the hydrogen samples. The method adopted was to apply a constant voltage to the filament and measure the resistance by means of a standard potentiometer circuit. This method has already been described by Bolland & Melville (1936).

Since some observers (Olson & Myers 1926; Moore & Taylor 1940) indicate the formation of significant amounts of methane as a result of the interaction of hydrogen atoms with olefines, it was considered necessary to investigate the effect of small amounts of methane on the thermal conductivity of a mixture with hydrogen.

The gauge was calibrated by making up known mixtures of normal hydrogen and methane, and the curves obtained were compared with those obtained by calculation. Since the prime interest was the measurement of the concentration of para-hydrogen in presence of methane, the determination was really that of a ternary mixture. There were two possible methods of dealing with this particular problem. The first is that described by Bolland & Melville (1936), who observed the voltage necessary to maintain the filament at two different temperatures, and the second is that described by Birse & Melville (1940*a*), involving two measurements only at room temperature. One measurement was made with the sample removed from the reaction vessel and the other with the same sample which had been allowed to come to equilibrium catalytically on the surface of the heated platinum spiral already mentioned.

The method adopted in the measurement was as follows. Mixtures of para-hydrogen and methane were made, containing up to 20 % methane, and were introduced to the gauge and adjusted to 50 mm. pressure. The resistance of the filament was determined and the mercury then lowered to expose the mixture to the glowing platinum equilibrator until all the para-hydrogen was converted to normal hydrogen and the resistance of the filament again determined when the pressure had been readjusted. The graph of gauge resistance against the percentage of methane present in the mixture is shown in figure 8. The upper full curve is that for a methane-normal hydrogen mixture, and the lower full curve for a methane-para-hydrogen mixture (36 % *p*-hydrogen).

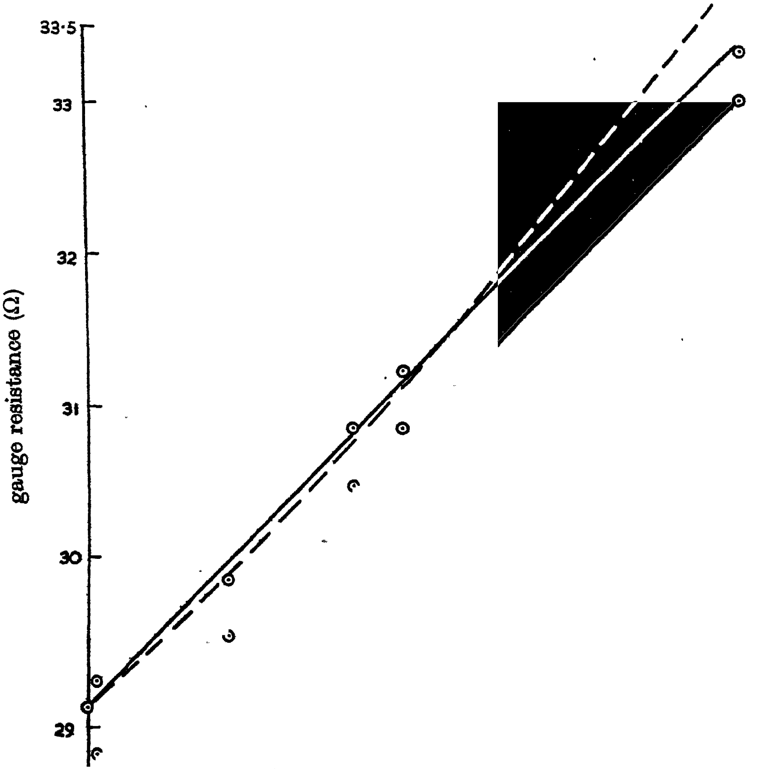


FIGURE 8. Effect of methane on thermal conductivity of hydrogen.

The curve obtained by calculation for methane-normal hydrogen mixtures is shown by the broken line. It was constructed as follows.

For a gauge containing a gas and being operated under constant voltage, variable resistance conditions, the heat input to the spiral is given by

$$Q = E^2/R. \tag{1}$$

Now, this heat is being conducted away by the gas and

$$Q' = BK_g, \tag{2}$$

where B is a constant for a particular gauge operating at constant wall temperature depending mainly on the geometry of the system and K_g is the thermal conductivity of the gas. Since the loss of heat by conduction through the leads attached to the gauge is negligible (Bolland & Melville 1936), then equations (1) and (2) can be equated:

$$E^2/R = BK_g, \quad A/R = K_g, \tag{3}$$

where A is a constant defined by E^2/B.

For the particular gauge in use, the wall temperature was 90° K and the filament temperature was 210° C, i.e. the average temperature of the gas was 150° K. With pure normal hydrogen in the gauge, the resistance of the filament was found to be 28·875Ω. The thermal conductivity of normal hydrogen at 150° K is given as $0{\cdot}231 \times 10^{-3}$ cal./cm.sec.degree (Kennard 1938).

From equation (3),

$$A = R \times K_{H_2}^{150^\circ K} = 6{\cdot}665 \times 10^{-3}.$$

This factor will now be referred to as the gauge factor.

In presence of a fraction n of methane, assuming that the mixture law holds,

$$\left.\begin{aligned} A/R_1 &= K_{H_2}(1-n)+K_{CH_4}n \\ &= K_{H_2}+n(K_{CH_4}-K_{H_2}), \\ 1/R_1 &= K_{H_2}/A+n(K_{CH_4}-K_{H_2})/A, \end{aligned}\right\} \qquad (4)$$

where R_1 is the resistance of the gauge spiral in presence of a fraction n of methane.

The thermal conductivity of methane is given as $0{\cdot}037 \times 10^{-3}$ cal./cm.sec.degree approximately, therefore,

$$K_{H_2}/A = 0{\cdot}0346, \quad K_{CH_4}/A - K_{H_2}/A = -0{\cdot}0293.$$

Table 1 gives the calculated values of R_1 in the specified range.

TABLE 1

% CH_4	n	$n(K_{CH_4}-K_{H_2})/A$	$1/R_1$	R_1
0·0	0·0	0·0	0·0346	28·90
5·0	1/20	−0·00147	0·0333	30·20
10·0	1/10	−0·00293	0·0317	31·55
16·6	1/6	−0·00490	0·0297	33·65
20·0	1/5	−0·00586	0·0288	34·70

Reference to figure 8 shows that the agreement holds good up to about 10 % methane, and that above this figure the curves separate. It is suggested that this may be due to the fact that the simple mixture law is not strictly valid at these higher concentrations or that the rate of mixing of the gases is not fast enough.

The slope of the curve can be calculated as follows: Differentiating equation (4) with respect to n,

$$-A/R_1^2\, dR/dn = -K_{H_2}+K_{CH_4}, \quad dR/dn = R_1^2(K_{H_2}-K_{CH_4})/A. \qquad (5)$$

From equation (5), the slope of the curve can be calculated. For 5 % methane content, resistance 30·20 Ω,

$$dR/dn = 2{\cdot}66\Omega.$$

From the experimental curve, the observed increase is 2·75Ω, which is considered satisfactory agreement in view of the lack of precise data on the actual thermal conductivities of the gases used.

The next point to be elucidated is the effect of the presence of methane on the change of resistance of the filament when a sample containing para-hydrogen is equilibrated.

For a gauge containing a mixture of para-hydrogen, and ortho-hydrogen at constant operating voltage,

$$A/R_1 = K_p d + K_o(1-d), \tag{6}$$

where d is the fraction of para-hydrogen in the mixture, K_p the thermal conductivity of para-hydrogen, K_o the thermal conductivity of ortho-hydrogen.

For normal hydrogen

$$A/R_2 = K_n. \tag{7}$$

In presence of a fraction of n of methane,

$$A/R_3 = K_{\mathrm{CH}_4} n + (1-n)\{K_p d + K_o(1-d)\}, \tag{8}$$

where para-rich hydrogen is the other constituent of the mixture.

For a fraction n of methane in normal hydrogen,

$$A/R_4 = K_{\mathrm{CH}_4} n + (1-n) K_n, \tag{9}$$

where R_3 is the resistance of the filament in presence of para-hydrogen, ortho-hydrogen and methane, R_4 is the resistance in presence of normal hydrogen and methane, and K_n is the thermal conductivity of normal hydrogen.

From equations (6) and (7),

$$1/R_2 - 1/R_1 = K_n/A - \{K_p d + K_o(1-d)\}/A. \tag{10}$$

From equations (8) and (9),

$$\begin{aligned} 1/R_4 - 1/R_3 &= \{K_{\mathrm{CH}_4} n + (1-n) K_n - K_{\mathrm{CH}_4} n - (1-n)(K_p d + K_o(1-d))\}/A \\ &= \{K_n(1-n) - (1-n)(K_p d + K_o(1-d))\}/A \\ &= \{(1-n)(K_n - K_p d - K_o(1-d))\}/A \\ &= (1-n)(1/R_2 - 1/R_1). \end{aligned} \tag{11}$$

Therefore $$(R_4 - R_3)/R_3 R_4 = (1-n)(R_2 - R_1)/R_1 R_2. \tag{12}$$

Now, if small changes in resistance are considered due to the presence of small quantities of methane, equation (12) can be simplified approximately to

$$(R_4 - R_3)/(R_2 - R_1) = (1-n) R_4^2/R_2^2, \tag{13}$$

since $R_3 \doteqdot R_4$ and $R_1 \doteqdot R_2$.

Now, for small concentrations of methane, say 5 %, a concentration not to be exceeded, if ever reached, substituting in (13),

$$(R_4 - R_3)/(R_2 - R_1) = 0{\cdot}95 \times 30{\cdot}20^2/28{\cdot}90^2 = 1{\cdot}04. \tag{14}$$

That is to say, a change in the difference of the resistances for para-hydrogen and normal hydrogen of 4 % occurs when there is methane present to the extent of 5 %. The error involved can then be disregarded, since this concentration of methane will never be obtained.

Hence, in the determination of the percentage of para-hydrogen in a given sample, by determining the resistance change dR_1 when a sample of known para-content P % is equilibrated and then taking the sample with methane and measuring the corresponding change in resistance dR_2 on equilibration, the para-concentration can be determined by the formula,

$$P' - 25 = (P - 25)\, dR_2/dR_1,$$

where P' is the unknown para concentration, provided the change dR is small compared with the value of R.

Measurement of the 2537 Å *quantum input to the reaction vessel*

This was accomplished by the method described by Birse & Melville (1940*b*) by making use of the conversion of para-hydrogen by hydrogen atoms generated by the mercury photo-sensitized dissociation of molecular hydrogen.

It was shown that the curve obtained by plotting the absolute rate of conversion of para-hydrogen as a function of pressure became pressure independent in a region of pressure at about 1 mm. In this part of the curve, conversion occurs exclusively by the mechanism

$$\mathrm{Hg} + h\nu = \mathrm{Hg}^*, \quad \mathrm{Hg}^* + p\text{-}\mathrm{H}_2 = \mathrm{Hg} + \mathrm{H} + \mathrm{H}, \quad \mathrm{H} = \tfrac{1}{2}\mathrm{H}_2 \text{ on the walls.}$$

Hence in this region, each quantum absorbed in the system results in the conversion of one molecule of hydrogen. Since the reaction is exclusively dissociation and recombination, the rate of production of hydrogen atoms is

$$d\mathrm{H}/dt = 2(\text{rate of dissociation of } p\text{-}\mathrm{H}_2) \times 100/\%\, p\text{-}\mathrm{H}_2 \text{ in gas},$$

allowing for dissociation of ortho molecules also, i.e.

$$d\mathrm{H}/dt = -(200/u)\,(d(p\text{-}\mathrm{H}_2)/dt), \tag{15}$$

where u is the percentage para-hydrogen in the original sample, or

$$d(\mathrm{H})/dt = -\frac{200}{u} \times \frac{d(p\text{-}\mathrm{H}_2)}{dt} \times \frac{1}{760} \times \frac{273}{T} \times \frac{1}{22{\cdot}4 \times 60} \text{ mole/l./sec.},$$

$$d(\mathrm{H})/dt = -\frac{200 d(p\text{-}\mathrm{H}_2)/dt}{u \times 760} \times \frac{273}{T} \times \frac{6{\cdot}06 \times 10^{23}}{22{\cdot}4 \times 10^{3}} \times \frac{V}{60} \text{ atom/sec.},$$

in the whole volume of the reaction vessel, where V is the volume of the reaction vessel in ml. The quantum input is obtained by dividing this number by 2, since each quantum produces two hydrogen atoms.

Table 2 illustrates one of such determinations. The values obtained for the absolute conversion rate can be extrapolated as in figure 9. The temperature of the

TABLE 2. QUANTUM INPUT DETERMINATION

pressure (mm.)	log p	exposure time (min.)	% para change	abs. conv. (mm.)	abs. conv. rate (mm./min.)
1·5	0·176	0·5	2·0	0·03	0·06
2·0	0·301	0·5	1·57	0·0314	0·063
3·0	0·477	1·0	2·36	0·0708	0·071
4·0	0·602	1·0	1·93	0·0772	0·077
6·0	0·778	1·0	1·47	0·0882	0·088
8·0	0·929	2·0	2·50	0·212	0·106
99·0	1·996	4·0	2·86	2·84	0·71

determination was 18° C, the volume of the reaction vessel 56·5 ml., the para-hydrogen content, u, 29 %, and the extrapolated value of the absolute conversion rate 0·057 mm./min. Substitution in the above formula gives the value of $0{\cdot}62 \times 10^{16}$ quanta/sec. for the quantum input to the reaction vessel.

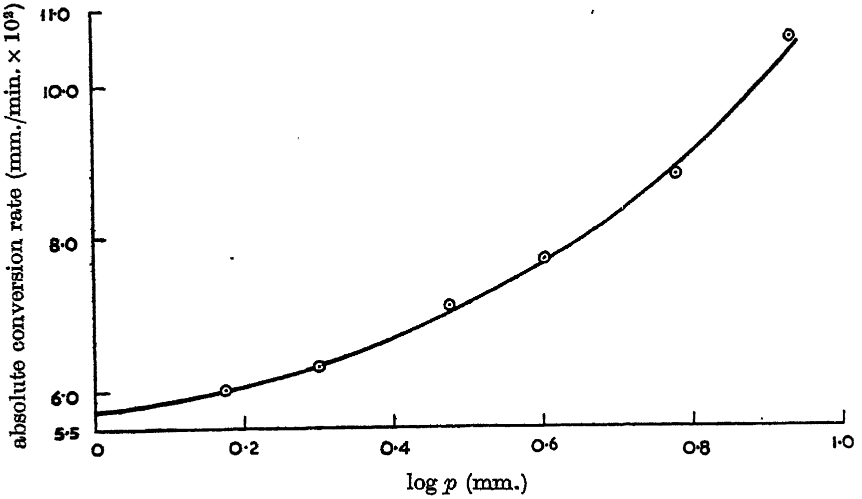


FIGURE 9. Effect of pressure on absolute rate of conversion of para-hydrogen.

Determination of the stationary concentration of atomic hydrogen

The stationary concentration of atomic hydrogen in a reaction vessel is defined by Birse & Melville (1940*b*)

$$(\mathrm{H}) = \frac{du/dt}{k^* u},$$

where du/dt is the rate of conversion of para-hydrogen by the exchange mechanism u is the para content initially and k^* is taken from the data of Geib & Harteck (1931) to be $1{\cdot}41 \times 10^5$.

Knowing $d(\mathrm{H})/dt$ and (H), the lifetime of a hydrogen atom can be computed, since

$$\tau = (\mathrm{H})/(d\mathrm{H}/dt).$$

From the data in table 2 for a pressure of 99 mm., the rate of conversion of para-hydrogen is 0·71 mm./min., and the rate of conversion due to dissociation and recombination is 0·057 mm./min., therefore the rate of conversion due to exchange is 0·65 mm./min.

Substitution in the above formula for (H) gives

$$(\mathrm{H}) = 2{\cdot}32 \times 10^{-9}\,\mathrm{mole/l.}$$

and

$$\tau = 1{\cdot}13 \times 10^{-1}\,\mathrm{sec.}$$

Conversion of para-hydrogen in presence of ethylene

It was hoped in the early stages of this work to use the reaction

$$\mathrm{H} + p\text{-}\mathrm{H}_2 = n\text{-}\mathrm{H}_2 + \mathrm{H},$$

to measure the stationary hydrogen atom concentration in a mixture of hydrogen atoms, hydrogen molecules and ethylene. By then measuring the rate of removal of ethylene and knowing the concentration of atomic hydrogen and ethylene, the velocity constant k could have been evaluated:

$$-d(\mathrm{Et})/dt = k(\mathrm{H})(\mathrm{Et}).$$

It can be shown that by plotting the logarithm of the percentage of toal conversion against time or irradiation, a straight line will be obtained, and that this slope will be proportional to the stationary hydrogen atom concentration. Hence by carrying out a run with only para-hydrogen and then with para-hydrogen and ethylene, the stationary concentration of atomic hydrogen can be evaluated as already described, and the ratio of the slope in presence of ethylene to the slope in absence of ethylene will give the ratio in which the stationary concentration of atomic hydrogen has been reduced.

In figure 10 the result of such an experiment has been shown. It is seen that in curve B, an inhibition period is present followed by a reduced rate of conversion. This is due to the very efficient removal of all hydrogen atoms by the ethylene, completely preventing the conversion of para-hydrogen. The reduced rate results from the reduction in the stationary concentration of hydrogen atoms due to reaction between the hydrogen atoms and the paraffin products.

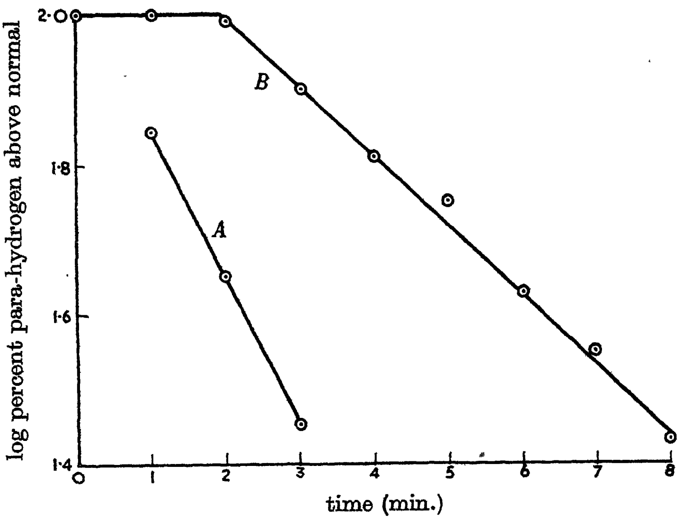


FIGURE 10. Inhibition of conversion of para-hydrogen by ethylene. Curve A, 100 mm. hydrogen. Curve B, $+0{\cdot}65$ mm C_2H_4.

As a matter of interest, two runs were done with n-hexane and *cyclo*hexane to determine whether or not the whole of the induction period could be attributed to the olefine. The data obtained are plotted in figure 11. It is seen that, qualitatively, n-hexane reduces the stationary concentration of atomic hydrogen more than *cyclo*hexane and hence reacts more readily with a hydrogen atom. From the data presented here, the collision efficiency of the interaction of hydrogen atom with these two hydrocarbons can be obtained.

From the curves in figure 11 it is seen that the reduction in the stationary concentration of hydrogen atoms (and hence the life-time of the atom) in the case of *cyclo*hexane is by a factor of 2·34. The lifetime of a hydrogen atom in 99 mm. hydrogen has been already calculated at $1{\cdot}13\times10^{-1}$ sec., so that in presence of 10 mm. *cyclo*hexane, this figure is reduced to

$$1{\cdot}13\times10^{-1}/2{\cdot}34\,\text{sec.} = 4{\cdot}85\times10^{-2}\,\text{sec.}$$

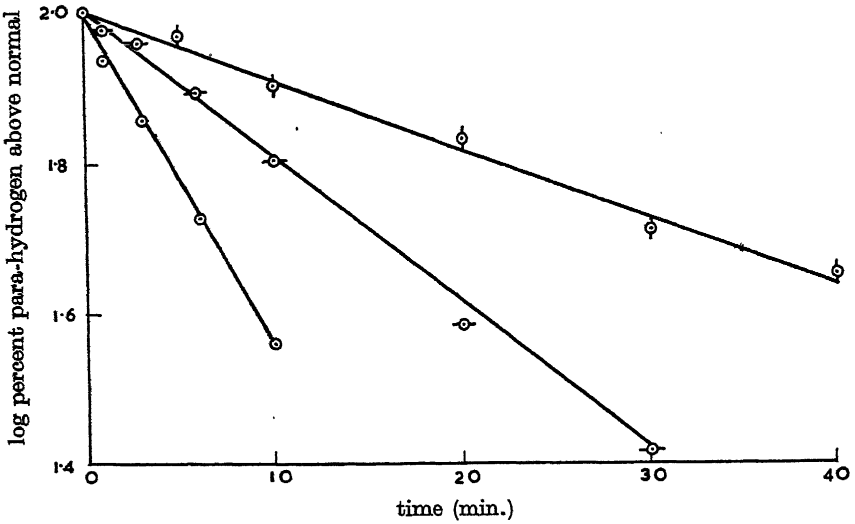


FIGURE 11. Effect of normal and *cyclo*hexane on conversion of para-hydrogen.
⊙ 100 mm. *p*-hydrogen. ϕ +11·5 mm. *n*-hexane. -⊙- +10 mm. *cyclo*hexane.

From the relationship

$$1/\lambda = \pi N_2(d_1/2+d_2/2)^2\{(M_1+M_2)/M_2\}^{\frac{1}{2}},$$

where λ is the mean free path of a molecule 1 meeting another molecule 2, the mean free path for a hydrogen atom between collisions with *cyclo*hexane molecules can be obtained. N_2 is the number of molecules of type 2 present in 1 ml. of gas, d_1 is the cross-section of a hydrogen atom and d_2 is that of the hydrocarbon molecule.

For 10 mm. *cyclo*hexane,

$N_2 = 3{\cdot}56\times10^{17}$ mol./ml.,

$d_1 = 2{\cdot}74\times10^{-8}$ cm. (Bonhoeffer & Harteck 1933),

$d_2 = 7\times10^{-8}$ cm. by measurement of atomic models,

From the above formula, λ is computed to be $3{\cdot}7\times10^{-4}$ cm.

The velocity of a hydrogen atom can be taken at $2{\cdot}4\times10^{5}$ cm./sec., therefore the number of collisions between a hydrogen atom and a *cyclo*hexane molecule will be $2{\cdot}4\times10^{5}/3{\cdot}7\times10^{-4} = 6{\cdot}5\times10^{8}$ collisions/sec.

The average lifetime of a hydrogen atom in this gas mixture has been shown to be $4{\cdot}85\times10^{-2}$ sec., therefore the number of collisions made between the average hydrogen atom and *cyclo*hexane molecules before removal of the hydrogen atom by reaction will be $6{\cdot}5 \times 10^8 \times 4{\cdot}85 \times 10^{-2}$ collisions, i.e. $3{\cdot}16 \times 10^7$ collisions. Hence the collision efficiency of the reaction is $1/3{\cdot}16 \times 10^7 = 3{\cdot}16 \times 10^{-8}$.

In the same way, for normal hexane, the lifetime of the hydrogen atom is reduced by a factor of 5·0. This gives a lifetime of $2{\cdot}3 \times 10^{-2}$ sec.

$$N_2 = 4{\cdot}1 \times 10^7 \text{ mol./ml.},$$

$$d_1 = 2{\cdot}74 \times 10^{-8} \text{ cm.},$$

$$d_2 = 9 \times 10^{-8} \text{ cm. by measurement of atomic models.}$$

λ is computed to be $2{\cdot}2 \times 10^{-4}$ cm., and hence the number of collisions between a hydrogen atom and a *n*-hexane molecule is $1{\cdot}1 \times 10^9$ collisions/sec. The average number of collisions before removal will be $2{\cdot}5 \times 10^7$ collisions. Hence the collision efficiency of the reaction is $4{\cdot}0 \times 10^{-8}$.

These two examples have been included in order to illustrate how the parahydrogen conversion technique can be applied to an investigation of reactions whose collision efficiencies are comparable with that of the reaction

$$H + p\text{-}H_2 = n\text{-}H_2 + H.$$

The fact that the para conversion is entirely inhibited by small quantities of ethylene renders the method useless as a means of determining the collision efficiency of the reaction $H + C_2H_4 = C_2H_5$.

A new competitive reaction had to be sought and the removal of hydrogen atoms on the white oxide of molybdenum or the yellow trioxide of tungsten was finally adopted. There were two points which had to be investigated with regard to the removal of hydrogen atoms on molybdenum oxide. It was necessary to establish (i) that in the reaction vessel every hydrogen atom produced was removed on the oxide; (ii) that the hydrogen atom was removed on only one collision with the oxide surface.

The first point was decided by measuring the quantum input to the reaction vessel by means of the para conversion and comparing this value with that obtained by measuring the number of hydrogen atoms removed by the oxide.

The quantum input to the reaction vessel has been already obtained above. This vessel had, however, a thin layer of polymer on the glass due to continued use, and by comparing the rate of para-conversion before and after removal of this layer, the quantum input was found to be cut down by a factor of 3·65 by the polymer. This gave a quantum input to the clean vessel of $2{\cdot}28 \times 10^{16}$ quanta-sec., therefore the rate of dissociation of molecular hydrogen was $2{\cdot}28 \times 10^{16}$ mol./sec. This corresponds to a volume of gas of 0·051 ml. at N.T.P./min.

By following the rate of removal of hydrogen in the same reaction vessel which had for this part of the experiment a layer of molybdenum oxide on the bottom, the rate of removal was averaged to be about 0·47 mm./min. The volume of the reaction vessel and manometer system was 96·7 ml., and the volume of hydrogen removed under these conditions was 0·057 ml. at N.T.P./min. The agreement between these estimates is sufficient to show that removal is quantitative. The thickness of the reaction vessel was 20 mm.

It had now to be shown that only one collision was necessary for removal of the hydrogen atom. The para-hydrogen technique was also used in this investigation. The same reaction vessel as that in the previous experiments was used, and the lower internal surface was coated with molybdenum oxide.

If the assumption is made that every collision of a hydrogen atom with the removing layer is effective in removing the atom, the average path distance of the atom can be calculated, and hence the number of collisions of a hydrogen atom with para-hydrogen molecules can be obtained. By following the para conversion, the number of fruitful collisions can be obtained and the ratio of these two gives the collision efficiency of the reaction

$$H + p\text{-}H_2 = n\text{-}H_2 + H.$$

This value of the collision efficiency can be compared with the values obtained by previous workers, and if the two estimates agree, then the assumption regarding the absolute efficiency of the addition of a hydrogen atom to the molybdenum trioxide is justified.

If there are N molecules in the reaction vessel and if the percentage of para-hydrogen is k, then there are $Nd/100$ para-hydrogen molecules in the vessel. After irradiation, suppose the quantum input is Q, then the loss of molecules due to decomposition and removal on the oxide will be Qt, where t is the time of irradiation and the nett number of molecules left will be $N-Qt$. Since in the decomposition, the ratio of the number of para-hydrogen molecules decomposed to the number of ortho-hydrogen molecules decomposed is in the percentage ratios of the individual components, then no change in the para percentage will occur as a result of the initial stage of dissociation.

If, in its path through the reaction vessel, each atom makes n fruitful collisions, then the total number of collisions will be $2nQt$. The total number of 'H_3' complexes formed will be $2nQt$, and since the exchange will proceed in the ratio of 1 para-molecule for every four complexes broken up, there will be $(2nQt \times 25)/100$ para-hydrogen molecules regenerated and since in the formation of the complex 'H_3', $(2nQtd/100)$ para-hydrogen molecules have been removed, the nett loss of para-hydrogen molecules is

$$2nQt(d-25)/100.$$

The new percentage of para-hydrogen will then be

$$d_1 = \{Nd - 2nQt(d-25)\}/(N-Qt).$$

By measuring the quantum input to the reaction vessel by the method described above, the initial fraction of para-hydrogen and the fraction of para-hydrogen after a certain time of irradiation, the number n can be determined.

The results of three runs of this type are shown in table 3. The quantum input to the reaction vessel was $2{\cdot}92 \times 10^{16}$ quanta/sec.

The total number of collisions which each atom makes with hydrogen molecules is obtained as follows.

The diffusion coefficient for hydrogen atoms in hydrogen gas at a hydrogen pressure of 10 mm. is 80 cm.^{2}/sec. Since $x^2 = 2Dt$, where x is the distance travelled by the atom in time t, and in the reaction vessel, the thickness is 2·0 cm., then $t = 0{\cdot}025$ sec. That is, the average lifetime of a hydrogen atom in the reaction vessel is 0·025 sec. Since the velocity of a hydrogen atom is $2{\cdot}4 \times 10^5$ cm./sec., its total path length will be, in 0·025 sec., $6{\cdot}0 \times 10^3$ cm.

The mean free path is given by the formula as already stated

$$1/\lambda = \pi N_2(d_1/2+d_2/2)^2\{(M_1+M_2)/M_2\}^{\frac{1}{2}} = 1030.$$

That is, for the diffusion of hydrogen atoms through molecular hydrogen at 10 mm. pressure, $\lambda = 1/1030$ cm.

Table 3

run number	pressure (H_2 mm.)	% p-H_2 initial	% p-H_2 final	time of irradiation (sec.)	N (molecules) ($\times 10^{19}$)	number of fruitful collisions	total number of collisions/atom ($\times 10^6$)	collision efficiency ($\times 10^{-7}$)
1	10	36·0	34·73	30	2·01	2·86	6·18	4·6
2	10	36·0	34·2	30	2·01	3·63	6·18	5·8
3	10	36·0	33·4	60	2·01	2·98	6·18	4·7

The number of collisions made by a hydrogen atom in its path through the reaction vessel will be given by dividing the total path distance by the mean free path, i.e. the number of collisions made by each hydrogen atom will be

$$6{\cdot}0 \times 10^{13} \times 1030 = 6{\cdot}18 \times 10^6.$$

By dividing the number of effective collisions per atom by the total number of collisions per atom, the collision efficiency of the exchange reaction of a hydrogen atom with a molecule of hydrogen can be obtained. These results are also shown in table 3.

The values obtained by Geib & Harteck (1931) are around 4×10^{-7}, which is in agreement with the estimates here.

Further confirmation of the absolute efficiency of the removal of hydrogen atoms on molybdenum oxide is provided by Johnson (1928), who used molybdenum trioxide to measure the angles of reflexion of beams of atomic hydrogen from crystals. The precision of measurement of the beam angle is entirely dependent on the sharpness of the mark on the oxide plate. We have observed also the sharpness of the 'shadow' effect obtained by irradiating hydrogen in a limited region above a surface of molybdenum trioxide.

References

Birse, E. A. B. & Melville, H. W. 1940*a* *Proc. Roy. Soc.* A, **175**, 164.
Birse, E. A. B. & Melville, H. W. 1940*b* *Proc. Roy. Soc.* A, **175**, 187.
Bolland, J. L. & Melville, H. W. 1936 *Trans. Faraday Soc.* **32**, 1525.
Bonhoeffer, K. F. & Harteck, P. 1933 *Grundlagen der Photochemie.* Leipzig: Steinkopff.
Geib, K. H. & Harteck, P. 1931 *Z. Phys. Chem.* p. 849.
Johnson, T. H. 1928 *J. Franklin Inst.* **206**, 301.
Kennard, E. H. 1938 *Kinetic theory of gases.* New York: McGraw-Hill.
Liebert, M. 1931 *Mellor's Inorganic Chemistry,* XI, 536. London: Longmans, Green & Co.
Moore, W. J. & Taylor, H. S. 1940 *J. Chem. Phys.* **8**, 504.
Olson, A. R.. & Myers, C. H. 1926 *J. Amer. Chem. Soc.* **48**, 389.

The kinetics of the interaction of atomic hydrogen with olefines

II. Diffusion theory

By H. W. Melville, F.R.S. and J. C. Robb

Chemistry Department, Marischal College, Aberdeen

(*Received* 23 *June* 1948—*Read* 10 *February* 1949)

In this paper, a system is described in which diffusion of a hydrogen atom takes place effectively in one dimension. The exact differential diffusion equations can be set up and solved, taking into account the possibility of gas-phase removal of the hydrogen atom by a reaction such as $H + C_2H_4 = C_2H_5$.

The collision efficiency of such a reaction has been related to the fraction of hydrogen atoms which reach the oxide layer under certain well-defined conditions. The calculated distribution curves of hydrogen atom concentration throughout the reaction vessel are also given under various conditions.

Introduction

In part I it has been shown that in the presence of an olefine, hydrogen atoms are so rapidly removed that the stationary state concentration is virtually zero. Certainly, the concentration of hydrogen atoms produced by the mercury-photosensitized dissociation of molecular hydrogen is so low as to be not measurable by the para-hydrogen technique. Hence it was necessary to find some other technique for the controlled removal of hydrogen atoms at a rate which could compete on equal terms with the olefine.

The technique adopted was to place a layer of white molybdenum oxide in such a position that a hydrogen atom had a roughly equal chance of reaching the layer or being removed by the olefine. Thus by placing the layer parallel to a silica window in a reaction vessel, through which light of wave-length 2537 Å was passed, a concentration gradient of hydrogen atoms could be established due to migration of the atoms from their point of generation to the removing layer. Introduction of olefine then disturbed this gradient and reduced the number of hydrogen atoms which could reach the removing surface. The problem was then to relate the efficiency of the interaction of the atom with the olefine to the change in concentration gradient and hence to the number of atoms which were able to survive olefine collisions and reach the removing surface.

Diffusion theory

It is shown elsewhere that the efficiency of interaction of a hydogen atom with molybdenum oxide is absolute.

Consider a hydrogen atom generated in the reaction vessel in presence of a certain pressure of an olefine. The atom has two possible fates: (1) it can diffuse to the oxide and be removed, or (2) it can be removed by the olefine as the result of a fruitful collision. The system is then that of a particle diffusing in one dimension and being subjected to the possibility of removal by an olefine molecule before reaching its target on the oxide layer. It is thus necessary to solve the exact diffusion equations and to take due account of the possibility of gas-phase removal of the hydrogen atom.

By obtaining a theoretical relationship between the fraction of hydrogen atoms which reach the oxide and the efficiency of the gas-phase removal, the latter factor can be obtained from an experimental determination of the former.

THE DIFFUSION COEFFICIENT

The diffusion coefficient is defined as the amount of material passing in unit time across a plane of 1 sq.cm. when the concentration gradient, $-dn/dx$, is unity. This gives the following formula:

$$x^2 = 2Dt,$$

where x is the distance travelled in time t and the diffusion coefficient is D.

In a calculation of the time for an atom or molecule to diffuse through a certain distance it is not possible to obtain precise data, since the calculation of the diffusion coefficient, especially in a mixture of gases, is attended by a considerable approximation. For example, the formula for calculating D differs by a numerical factor according to Meyer and Smoluchowsky. Meyer's (1899) estimate is

$$D = \eta/\rho = \tfrac{1}{3}(\bar{u}l),$$

and that of Smoluchowsky (1906) is

$$D = \frac{2}{3\pi}(\bar{u}l).$$

It is seen that the two estimates differ only in the numerical coefficient, and by adopting one of them at least a comparison of results can be obtained. In the above formulae, η is the viscosity of the medium, ρ is the density, $\bar{u}$ the average velocity of the particle and l the mean free path. The formula of Mayer will be adopted in the following treatments.

A further source of inaccuracy is the difficulty of obtaining reliable data on the molecular diameters, since the methods adopted in obtaining the information modify the results (Farkas & Melville 1939). The best estimates for this investigation are those obtained from viscosity measurements and these have been used wherever possible.

DETERMINATION OF THE DIFFUSION COEFFICIENT

For a system containing two species of molecule, 1 and 2, at a concentration respectively of N_1 and N_2 mol./c.c., and whose molecular diameters are respectively d_1 and d_2, then when l_1 is the mean free path for molecule 1,

$$1/l_1 = \pi\sqrt{2}\,N_1 d_1^2 + \pi N_2(\tfrac{1}{2}d_1 + \tfrac{1}{2}d_2)^2\sqrt{\{(M_1+M_2)/M_2\}}.$$

The first expression on the right-hand side represents the mean free path for molecule 1 meeting another molecule 1, and the second represents the mean free path for a molecule 1 meeting a molecule 2.

In the problem at hand, a system of three components is being dealt with; to take a specific case, hydrogen atoms, hydrogen molecules and ethylene molecules. Assuming that these molecules exert an additive effect, it is possible to extend the above formula to three components:

$$1/l_1 = \pi\sqrt{2}N_1d_1^2+\pi N_2(\tfrac{1}{2}d_1+\tfrac{1}{2}d_2)^2\sqrt{\{(M_1+M_2)/M_2\}}+\pi N_3(\tfrac{1}{2}d_1+\tfrac{1}{2}d_3)^2\sqrt{\{(M_1+M_3)/M_3\}}.$$

Since the concentration of hydrogen atoms will be low compared with the other species present, the first term on the right-hand side may be neglected. The subscript 1 refers to the hydrogen atoms, 2 to the hydrogen molecules and 3 to the ethylene molecules.

The diameter of the hydrogen atom is given as $2{\cdot}74\times10^{-8}$ cm. (Bonhoeffer & Harteck 1933), and the hydrogen molecule is also $2{\cdot}74\times10^{-8}$ cm. (Kennard 1938). The diameter of the ethylene molecule is $4{\cdot}95\times10^{-8}$ (Kennard 1938).

Substituting these values in the above formula,

$$1/l_1 = 28{\cdot}8\times10^{-16}N_2+47{\cdot}5\times10^{-16}N_3.$$

Now, in 1 ml. of gas at a pressure of 1 mm. and at normal temperature, there are $3{\cdot}56\times10^{+16}$ mol., and hence at a pressure of p mm., $3{\cdot}56p\times10^{+16}$ mol. Then, using this figure in the above,

$$1/l_1 = 102{\cdot}7p_2+169{\cdot}1p_3,$$

where p_2 and p_3 are respectively the pressures of hydrogen and ethylene.

Now, applying Meyer's formula,

$$D = \tfrac{1}{3}(\overline{u}_{\mathrm{H}}/(102{\cdot}7p_2+169{\cdot}1p_3)\ \text{cm.}^2/\text{sec.}$$

For a molecule, the average velocity $\overline{u}$ is given by

$$\overline{u} = k/M^{\frac{1}{2}},$$

where k is a constant and M the molecular weight of the species. For hydrogen molecules, the average velocity is $1{\cdot}692\times10^5$ cm./sec., therefore for a hydrogen atom, at normal temperature, it will be $2{\cdot}4\times10^5$ cm./sec.

$$D = 8\times10^4/(102{\cdot}7p_2+169{\cdot}1p_3)\ \text{cm.}^2/\text{sec.}$$

TABLE 1. DIFFUSION FORMULAE FOR THE CALCULATION OF D

Ethylene. Diameter of molecule = $4{\cdot}95\times10^{-8}$ cm.
$D = 8\times10^4/(102{\cdot}7p_2+169{\cdot}1p_3)$ cm.²/sec.

Propylene. Diameter of molecule = $6{\cdot}4\times10^{-8}$ cm.
$D = 8\times10^4/(102{\cdot}7p_2+238p_3)$ cm.²/sec.

iso-Butylene. *Diameter of molecule = $6{\cdot}5\times10^{-8}$ cm.
$D = 8\times10^4/(102{\cdot}7p_2+240p_3)$ cm.²/sec.

n-Pentene-2 (*cis*). *Diameter of molecule = $8{\cdot}0\times10^{-8}$ cm.
$D = 8\times10^4/(102{\cdot}7p_2+320p_3)$ cm.²/sec.

2-3-3-*Trimethyl-butene*-1
*Diameter of molecule = $7{\cdot}5\times10^{-8}$ cm.
$D = 8\times10^4/(102{\cdot}7p_2+290p_3)$

* These estimates of the molecular diameters were made by constructing the appropriate molecules as Stuart models and measuring the effective diameters. This will not involve a large source of error, since p_3 is much less than p_2.

Table 1 gives the values of the cross-sections of the various molecules studied in this work and also the formulae for calculating the diffusion constant D for the various mixtures used.

PRECISE CALCULATION OF STATIONARY HYDROGEN ATOM CONCENTRATIONS

Semenoff (1935) has considered those chain reactions in which one active centre is continually replaced by another, but the problem on hand concerns a particle which retains its own identity until its final destruction. This, however, does not affect the physical concepts and similar methods may be directly applied.

The conditions which are used in this particular problem are somewhat different to those of Semenoff's calculations. Here, a reaction vessel is used in which the active centres are produced by irradiation of one plane surface of the vessel and are removed on a layer of molybdenum or tungsten oxides. Since the top surface has been shown to be a good reflector for hydrogen atoms, then it simplifies the conception and the mathematical treatment by considering the mirror image of the reaction vessel to be placed at the generating wall and the wall removed, since to a hydrogen atom reflected from the wall, the effect is the same as if it were coming from another vessel of the same dimensions.

In figure 1 the planes at O and B represent the walls of the reaction vessel. The plane at O is the reflecting wall and the plane at B the removing wall. Now construct the mirror image of this cell on the plane O. The plane at O is now considered to be non-existent for the purposes of the following calculations. The mirror image of B is A.

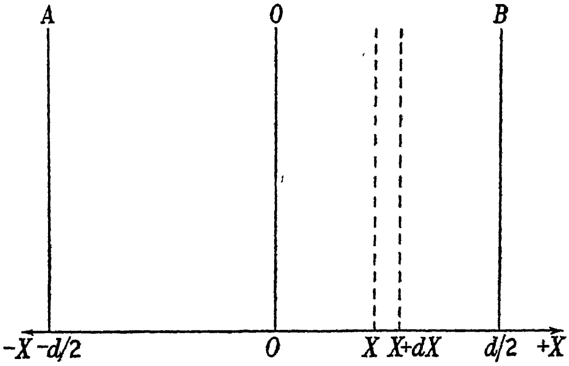


FIGURE 1. Diffusion system.

If the plane O is considered as the origin of co-ordinates, then the planes B and A are determined by the equations $x = \frac{1}{2}d$ and $x = -\frac{1}{2}d$ respectively.

The active centres are produced in the whole volume AB, and then diffusion to the walls takes place by virtue of the concentration gradient established in neighbouring elements of the vessel. On the walls A and B, the concentration of the active centres is zero and the maximum concentration of the active centres will be found in the centre of the vessel. On the plane O, the concentration gradient dn/dx will be zero.

Consider now two planes in the vessel, x and $x+dx$, separated by a very small distance dx, and let D denote the diffusion coefficient for the active centres through the gas mixture. Then the number of active centres which pass through the plane x in the direction of the wall will be $-D(dn/dx)_x$ per unit area. The number of centres passing through the plane $x+dx$ in the direction of the wall will similarly be given by $-D(dn/dx)_{x+dx}$. There are therefore $-D(dn/dx)_x$ centres which in unit time and area enter the element dx and

$$-D(dn/dx)_{x+dx} = -D(dn/dx)_x - D(d^2n/dx^2)_x\,dx$$

which leave it.

The absorption of radiation will follow the normal law. Where I is the intensity per square cm. at a point x cm. inside the reaction vessel, I_0 the intensity per sq.cm. of incident radiation, and k is the extinction coefficient, then

$$I = I_0\,\mathrm{e}^{-kx}. \tag{1}$$

Differentiating equation (1),

$$dI/dx = -kI_0\,\mathrm{e}^{-kx}.$$

That is, the amount of light absorbed at any particular point in the vessel is $kI_0\,\mathrm{e}^{-kx}$ quanta per unit volume. If each quantum of radiation produces c active centres, then at any point the number of active centres produced will be $kcI_0\,\mathrm{e}^{-kx}$.

Now considering removal solely on the walls, in the layer dx, $-D(dn/dx)_x$ centres enter and $-D(dn/dx)_{x+dx}$ leave, while $kcI_0\,\mathrm{e}^{-kx}\,dx$ centres are generated within the layer. Thus, in the stationary state,

$$-D\left(\frac{dn}{dx}\right)_x + kcI_0\,\mathrm{e}^{-kx}\,dx = -D\left(\frac{dn}{dx}\right)_{x+dx},$$

$$D\left(\frac{dn}{dx}\right)_{x+dx} - D\left(\frac{dn}{dx}\right)_x + kcI_0\,\mathrm{e}^{-kx}\,dx = 0,$$

i.e.

$$D\frac{d^2n}{dx^2} + kcI_0\,\mathrm{e}^{-kx} = 0,$$

where the first term represents the loss of centres due to diffusion and the other the generation of active centres due to adsorption of radiation.

Let $a = I_0kc/D$, then the equation becomes

$$d^2n/dx^2 = -a\,\mathrm{e}^{-kx}.$$

The solution of this type of equation is of the general form

$$n = A + Bx - a\,\mathrm{e}^{-kx}/k^2,$$

where A and B have to be determined. One of the boundary conditions is $n = 0$ at $x = \frac{1}{2}d$. From the symmetry of the system, n is a maximum at the centre, and since transport by diffusion would be equal in both directions, $dn/dx = 0$. This is the other condition.

Putting $n = 0$ at $x = \frac{1}{2}d$,

$$O = A + Bd/2 - a\,\mathrm{e}^{-kd/2}/k^2. \tag{2}$$

Putting $dn/dx = 0$ at $x = 0$ $\quad O = B + a/k. \tag{3}$

From equations (2) and (3), $\quad B = -a/k,$

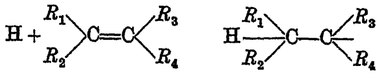
and $\quad A = a\,\mathrm{e}^{-kd/2}/k^2 + ad/2k.$

Hence
$$\left.\begin{aligned} n &= a\,\mathrm{e}^{-kd/2}/k^2+ad/2k-ax/k-a\,\mathrm{e}^{-kx}/k^2,\\ n &= a\{\mathrm{e}^{-kd/2}-\mathrm{e}^{-kx}+k(d/2-x)\}/k^2. \end{aligned}\right\} \tag{I}$$

In the case of removal of active centres in the gas phase by an olefine, the number of active centres removed in unit time will be given by

$$(n_r)_x = k_1(n_x)\,(X)_1,$$

where n_x is the stationary concentration of active centres in the layer, (X) is the concentration of the olefine molecules and k_1 the velocity constant of the reaction



To simplify the notation, put

$$g = k_1(X).$$

Then
$$(n_r)_x = g(n_x).$$

The differential equation obtained for this case is

$$D(d^2n/dx^2)+I_0kc\,\mathrm{e}^{-kx}-gn = 0.$$

Let
$$a = I_0kc/D, \quad b^2 = g/D.$$

Then the equation resolves to the form

$$d^2n/dx^2-b^2n = -a\,\mathrm{e}^{-kx},$$

where the first term accounts for loss of centres due to diffusion, the second for loss due to gas-phase reaction and the third generation of the active centres.

The solution of this equation for n will be of the following general form:

$$n = A\,\mathrm{e}^{bx}+B\,\mathrm{e}^{-bx}-(a/(k^2-b^2))\,\mathrm{e}^{-kx}. \tag{4}$$

As before, the boundary conditions are $n = 0$ at $x = \frac{1}{2}d$ and $dn/dx = 0$ at $x = 0$. Substituting these conditions in equation (4),

$$O = A\,\mathrm{e}^{bd/2}+B\,\mathrm{e}^{-bd/2}-(a/(k^2-b^2))\,\mathrm{e}^{-kd/2},$$

$$O = A-B+ak/b(k^2-b^2).$$

From these equations, A and B can be determined,

$$A = a(\mathrm{e}^{-kd/2}-k/b\,\mathrm{e}^{-bd/2})/2\cosh bd/2(k^2-b^2),$$

$$B = a(\mathrm{e}^{-kd/2}+k/b\,\mathrm{e}^{bd/2})/2\cosh bd/2(k^2-b^2).$$

Substituting these values for A and B in equation (4),

$$n = \frac{1}{k^2-b^2}a\left\{\frac{\mathrm{e}^{bx-kd/2}-k\,\mathrm{e}^{b(x-\frac{1}{2}d)}/b+\mathrm{e}^{-bx-kd/2}+k\,\mathrm{e}^{-b(x-\frac{1}{2}d)}/b}{2\cosh bd/2}-\mathrm{e}^{-kx}\right\}$$

$$= a\left\{\frac{\mathrm{e}^{-kd/2}\cosh bx+[k\sinh b(\frac{1}{2}d-x)]/b}{\cosh bd/2}-\mathrm{e}^{-kx}\right\}\Big/(k^2-b^2). \tag{II}$$

OBSERVATIONS ON THE FOREGOING CALCULATIONS

First of all, it is of interest to see the shape of the distribution curve for the stationary concentration of the active centres.

In the particular case applied in the work contained here, the reaction vessel has a top surface of silica to allow passage of radiation of wave-length 2537 Å. This radiation produces mercury atoms raised from the $^1S_0$ state to the $^3P_1$ state. Since the pressure of hydrogen in the reaction vessel is of the same order as the pressure of neon in the mercury lamp producing the 2537 Å radiation, pressure broadening of the 2537 Å line emitted will be counterbalanced by a similar pressure broadening of the absorption line in the reaction vessel. This means that substantially the whole emitted line will be absorbed.

The excited atoms of mercury will then undergo collision with hydrogen molecules and produce two hydrogen atoms by the reaction

$$\mathrm{Hg}(^3P_1) + \mathrm{H_2} = \mathrm{H} + \mathrm{H} + \mathrm{Hg}(^1S_0)$$

or

$$\mathrm{Hg}(^3P_1) + \mathrm{H_2} = \mathrm{Hg.H} + \mathrm{H}, \quad \mathrm{Hg.H} = \mathrm{Hg} + \mathrm{H}.$$

Under the conditions of pressure employed, no mercury atoms can re-radiate their energy before undergoing reaction. The value of the factor c for the number of active centres produced per quantum absorbed is therefore 2.

The hydrogen atoms so produced then diffuse uniformly through the reaction vessel until they are removed by either of two reactions:

(1) successful collision with a double bond in the gas phase, or

(2) collision with a layer of molybdenum oxide on the bottom surface.

Under these conditions, generation of active centres is exponential throughout the reaction vessel, and the shape of the distribution curve will be dealt with.

This type of generation will be discussed with reference to a reaction vessel 2 cm. deep in which over 90 % of the incident radiation is removed within the first 5 mm. of the vessel, and with reference to a reaction vessel 2 mm. deep in which the intensity is reduced to only about 50 % in the journey through the whole vessel.

The two types of conditions lead to somewhat different distribution curves.

CALCULATIONS INVOLVING THE 2 CM. REACTION VESSEL

For mercury vapour at room temperature, 18° C, the vapour pressure is $1{\cdot}0 \times 10^{-3}$ mm. The absorption coefficient of mercury for 2537 Å radiation has been determined by Zemansky (Mitchell & Zemansky 1934; Zemansky 1931), who has shown that for low concentrations of mercury, the absorption coefficient varies nearly linearly with the number of atoms per ml. in the following manner:

$$k = 1{\cdot}38 \times 10^{-13} N,$$

where N is the number of atoms per ml.

Applying this figure in the usual formula,

$$I = I_0 \mathrm{e}^{-kx}$$

where x is the thickness of the layer through which the beam, incident at intensity I_0, is passing, and I is the intensity of the emergent beam, for a mercury pressure of 0·001 mm.

$$I = I_0\,\mathrm{e}^{-5\cdot 0x}.$$

k is then 5·0.

As has been pointed out, $c = 2$.

For a typical run done with 10 mm. pressure of molecular hydrogen, the value of the diffusion coefficient is of the order of $D = 10^2$ cm.²/sec.

The thickness, w, of the reaction vessel, is 2 cm.

For a gas-phase reaction removing hydrogen atoms by collision with an olefine at a pressure of 1 mm., energy of activation 4 kcal. and steric factor 0·1, g is of the order of 10^4.

I_0 is of the order of 10^{15} quanta/sec./sq.cm.

The shape of the curve of the concentration of 'quanta' removed at any point in the vessel is found as follows:

$$I = I_0\,\mathrm{e}^{-kx}, \quad (dI/dx) = -kI_0\,\mathrm{e}^{-kx}.$$

TABLE 2

x (cm.)	e^{-5x}	$-5\,\mathrm{e}^{-5x}$	$(dI/dx)/(dI/dx)_0$
0·0	1·000	−5·000	1·000
0·05	0·779	−3·895	0·779
0·1	0·607	−3·035	0·607
0·2	0·368	−1·840	0·368
0·4	0·135	−0·675	0·135
0·6	0·050	−0·250	0·050
1·0	0·007	−0·035	0·007
2·0	0·000	−0·000	0·000

Table 2 shows the rate of production of active centres in the reaction vessel at various points. The curve is plotted in figure 2*c*.

When the active centres are removed entirely by diffusion to the wall, formula (I) gives

$$n = I_0 c\{\mathrm{e}^{-kw} - \mathrm{e}^{-kx} + k(w-x)\}/kD,$$

where I_0 is 10^{-15}, c is 2, k is 5·0, w is 2 cm., D is 10^2 cm.²/sec. and n is the concentration of hydrogen atoms per ml. expressed as a pressure in mm. of mercury (table 3). These results are plotted in figure 2*b*.

TABLE 3

x (cm.)	n (mm. pressure)
0·0	$1{\cdot}00\times 10^{-3}$
0·2	$0{\cdot}96\times 10^{-3}$
0·5	$0{\cdot}82\times 10^{-3}$
1·0	$0{\cdot}55\times 10^{-3}$
1·5	$0{\cdot}28\times 10^{-3}$
2·0	0·0

In the case where removal occurs both in the gas phase and on the wall, formula (II) gives

$$n = \frac{I_0 kc}{D(k^2-b^2)}\left\{\frac{\cosh bx\,\mathrm{e}^{-kw}}{\cosh bw} + \frac{k\sinh b(w-x)}{\cosh bw} - \mathrm{e}^{-kx}\right\},$$

where $I_0 = 10^{15}$, $w = 2$ cm., $c = 2$, $D = 10^2$ cm.^{2}/sec. and $b = 10$, since $b = \sqrt{(g/D)}$. The value of b is about that encountered in the reaction being investigated here (table 4). These results also are shown in figure 2*a*.

TABLE 4

x (cm.)	n (mm. pressure)
0·0	$2{\cdot}0 \times 10^{-5}$
0·2	$1{\cdot}2 \times 10^{-5}$
0·5	$3{\cdot}2 \times 10^{-6}$
1·0	0·0
2·0	0·0

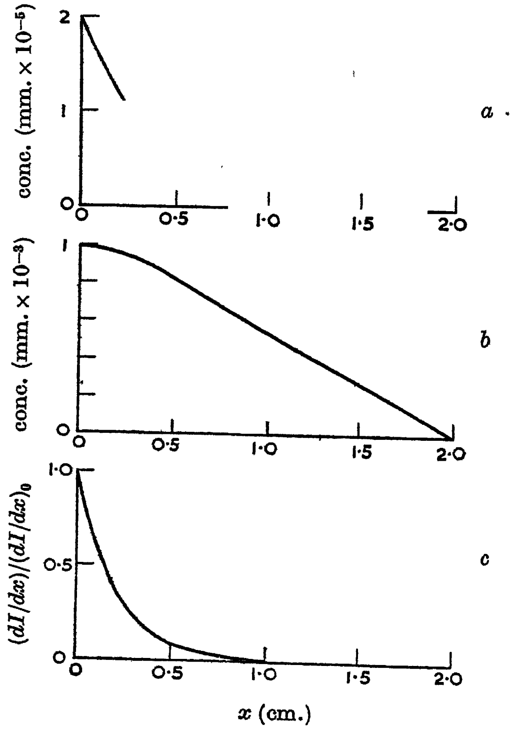


FIGURE 2. Rate of production of hydrogen atoms and the hydrogen atom distribution in the reaction vessel, 2 cm. thick. *a*, distribution of hydrogen atoms (gas-phase removal). *b*, distribution of hydrogen atoms (no gas-phase removal). *c*, absorption of 2537 Å radiation by mercury vapour at 18° C.

In the case of figure 2*a*, since the number of centres transported in the reaction vessel in unit time through a layer is proportional to the slope of the curve of concentration plotted against distance into the reaction vessel, the stationary concentration is so reduced that no appreciable number of hydrogen atoms will reach the removing layer. This means that in a reaction vessel of this dimension there is virtually no competitive reaction and the hydrogen atoms produced will all be removed in the gas phase. In the case of the reaction vessel 0·2 cm. thick, the position is quite different and some calculations involving this vessel will now be detailed.

CALCULATIONS INVOLVING THE 0·2 CM. REACTION VESSEL

The values of the various constants are as follows:

$$k = 5\cdot 0, \qquad c = 2, \qquad D = 10^2\,\text{cm.}^2/\text{sec.},$$
$$w = 0\cdot 2\,\text{cm.}, \qquad g = 10^4, \qquad I_0 = 10^{15}\,\text{quanta/sec.}$$

The curve relating amount of absorption per unit volume with the depth into the reaction vessel is given in figure 3*c*. The values used are those which are relevant from table 2. Removal by diffusion to the walls only requires the use of formula (I). Table 5 has been calculated from this formula. It may be noted that in this case, generation of active centres is proceeding throughout the whole volume of the reaction vessel. These results are plotted in figure 3*b*. From this graph it is seen that the distribution of hydrogen atoms in the gas phase does not now follow an exponential law. This is because the generation term is in this case important throughout the vessel, not just in the first fraction as with the 2 cm. vessel, indeed, the 2 cm. vessel case can be regarded almost as a narrow element generation.

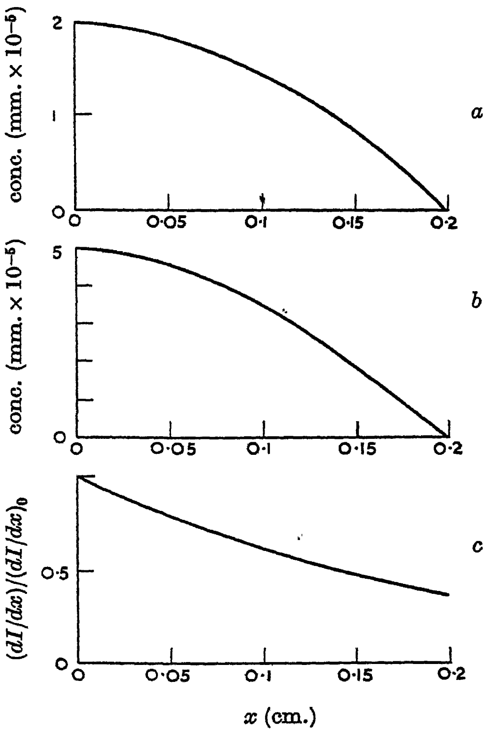


FIGURE 3. Rate of production of hydrogen atoms and the hydrogen atom distribution in the reaction vessel, 0·2 mm. thick. *a*, gas-phase removal. *b*, no gas-phase removal. *c*, absorption of 2537 Å.

For the case of removal in the gas phase as well as on the walls, formula (II) gives table 6. The graph of these results is shown in figure 3*a*. It will be seen that there is a finite slope at the wall of the reaction vessel which is responsible for the removal of hydrogen atoms. In this case, there is competition for the removal of hydrogen atoms between the gas-phase removing agent and the molybdenum oxide coated wall.

TABLE 5

x (cm.)	n (mm. pressure)
0·00	5×10^{-5}
0·02	$4{\cdot}9 \times 10^{-5}$
0·05	$4{\cdot}6 \times 10^{-5}$
0·10	$3{\cdot}5 \times 10^{-5}$
0·15	$1{\cdot}9 \times 10^{-5}$
0·20	0·0

TABLE 6

x (cm.)	n (mm.)
0·00	$2{\cdot}00 \times 10^{-5}$
0·02	$1{\cdot}97 \times 10^{-5}$
0·05	$1{\cdot}85 \times 10^{-5}$
0·10	$1{\cdot}43 \times 10^{-5}$
0·15	$0{\cdot}86 \times 10^{-5}$
0·20	0·00

APPLICATION OF THE PRECEDING DIFFUSION CALCULATIONS TO THE MEASUREMENT OF THE COLLISION EFFICIENCIES OF EFFICIENT GAS-PHASE REACTIONS

Since the most important case at the moment is that of the exponential production of active centres, it will be used throughout to illustrate the method of approach, although the same methods can be adopted for the cases of thin-layer generation and uniform generation.

The expression has already been formulated which gives the number of active centres present in unit volume of a particular element located at an average distance x from the plane of generation. It was

$$n = \frac{I_0 kc}{D(k^2-b^2)}\left\{\frac{\cosh bx\, e^{-kw}}{\cosh bw} + \frac{k \sinh b(w-x)}{b \cosh bw} - e^{-kx}\right\},$$

where w is the thickness of the reaction vessel.

Now, by definition, the diffusion coefficient D is given by

$$N' = -D(dn/dx)_x,$$

where N' is the actual number of active centres which pass through the plane x in unit time, $(dn/dx)_x$ is the concentration gradient at the point distance x from the plane of generation.

Differentiating the above expression for n and multiplying by D,

$$-N' = D(dn/dx)$$

$$-N' = \frac{I_0 kc}{(k^2-b^2)}\left\{\frac{b \sinh bx\, e^{-kw}}{\cosh bw} - \frac{k \cosh b(w-x)}{\cosh bw} + k\, e^{-kx}\right\},$$

where $b = \sqrt{(g/D)}$.

Consider the case of a reaction vessel where the top surface is the plane of entrance of the beam of light which generates the active centres and the lower is a plane where there is removal of every colliding centre.

If x in the equation just derived for N' is put equal to w, then, N'_w will be the number of active centres colliding with the absorbing layer and being removed. Further, in the equation, the terms $I_0 c$ can be eliminated as follows. The total amount of light which passes through the reaction vessel without being absorbed is

$$I = I_0 \, \mathrm{e}^{-kw},$$

so that the total amount of light absorbed by the mercury vapour in the reaction vessel will be

$$I_a = I_0(1 - \mathrm{e}^{-kw}).$$

The total number of active centres produced will then be $cI_0(1 - \mathrm{e}^{-kw})$. If the ratio of the number of active centres which reach the removing layer, N'_w, to the total number produced be R, then

$$R = N'_w / I_0 c(1 - \mathrm{e}^{-kw}).$$

Then substituting for $N'_w/I_0 c$ from the formula above,

$$R = \frac{k}{(1 - \mathrm{e}^{-kw})\,(k^2 - b^2)} \left\{ \frac{b \sinh bw \, \mathrm{e}^{-kw}}{\cosh bw} - \frac{k}{\cosh bw} + k\,\mathrm{e}^{-kw} \right\}.$$

In this equation, R is the observable quantity and b, the factor which is required. The only method of dealing with this equation is to set the initial conditions, i.e. pressure of mercury vapour, depth of vessel, etc., and then to plot values of R against chosen values of b. It is helpful if the approximate value of the diffusion coefficient and the collision efficiency are known. This enables a rough value to be calculated for b and the conditions chosen accordingly.

By suitable choice of conditions, a curve can be obtained to cover values of b over a factor of 10 and a change of R from say 0·25 to 0·75.

The following tables were prepared in this way. In table 7 is given the value of R for various values of b and for various vessel depths w. These values are plotted in figure 4, and it will be readily seen that the range of values of b which can be dealt with can be altered by changing the depth of the vessel.

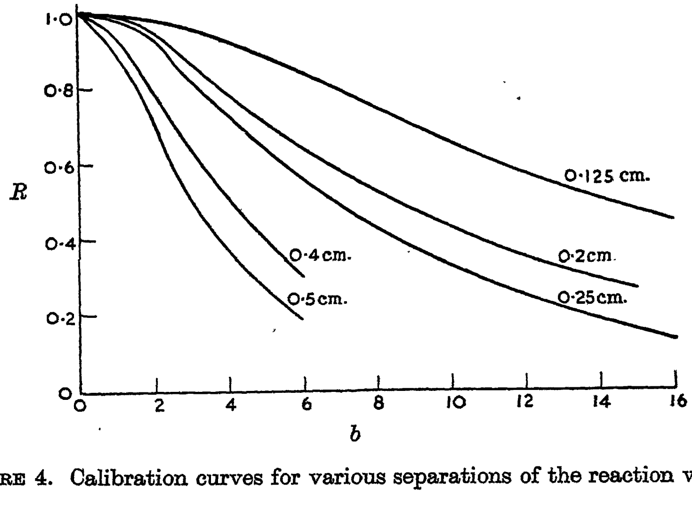


FIGURE 4. Calibration curves for various separations of the reaction vessel.

It was considered that the value of R might be extremely temperature dependent since the value of the mercury vapour pressure would change. To determine the extent of this effect, the following calculations were made.

For a change of temperature from 18 to 14° C, the mercury vapour pressure changes from 0·0010 to 0·0007 mm. and the corresponding change in k is from 5·0 to 3·5. The values of R were calculated for a plate separation of 0·25 mm. and for various values of b, using the new value for k. In table 8 the values of R for $k = 3{\cdot}5$ and $k = 5{\cdot}0$ are given. From this table, the close agreement in the values of R show that compensating factors render any close temperature control unnecessary.

TABLE 7

w (cm.)	0·125	0·20	0·25	0·40	0·50
b	R	R	R	R	R
0	1·00	1·00	1·00	1·00	1·00
1	—	0·99	—	0·94	0·90
2	—	0·95	0·94	0·80	0·70
3	—	—	0·82	0·63	0·51
4	0·92	0·78	0·73	0·50	0·37
6	0·84	0·66	0·56	0·30	0·19
8	0·75	0·54	0·42	—	—
10	0·66	0·43	0·32	—	—
12	0·58	0·35	0·26	—	—
15	—	0·27	0·16	—	—
16	0·45	—	—	—	—

TABLE 8

k	3·5	5·0
b	R	R
2	0·94	0·94
4	0·67	0·73
6	0·55	0·56
8	0·42	0·42
10	0·33	0·32

REFERENCES

Bonhoeffer, K. F. & Harteck, P. 1933 *Grundlagen der Photochemie*. Leipzig: Steinkopff.

Farkas, A. & Melville, H. W. 1939 *Experimental methods in gas reactions*. London: Macmillan & Co.

Kennard, E. H. 1938 *Kinetic theory of gases*. New York: McGraw-Hill.

Meyer, O. 1899 *Kinetic theory of gases*. London: Longmans, Green & Co.

Mitchell, A. C. G. & Zemansky, M. W. 1934 *Resonance radiation and excited atoms*. Cambridge University Press.

Semenoff, N. 1935 *Chemical kinetics and chain reactions*. Oxford: Clarendon Press.

Smoluchowsky, M. 1906 *Bull. Int. Acad. Cracovie*, p. 202.

Zemansky, M. W. 1931 *Phys. Rev.* **36**, 219.

The kinetics of the interaction of atomic hydrogen with olefines

III. Theory and technique involved in the use of the oxides of molybdenum and tungsten as hydrogen atom removers

By H. W. Melville, F.R.S. and J. C. Robb

Chemistry Department, Marischal College, Aberdeen

(*Received* 23 *June* 1948—*Read* 10 *February* 1949)

The colorimetric method of estimating the rate of addition of hydrogen atoms to the oxides of molybdenum and tungsten is discussed in detail. It is also shown that alkyl radicals are efficiently removed by molybdenum oxide, and allowance is made for the effect of their presence on the blueing rate of the oxide surface.

The method of evaluating collision efficiencies from the data obtained is indicated in full, and the construction and operation of a calculator to assist in the computation is described.

Introduction

It has been shown that the collision efficiency of a hydrogen atom with an olefine under the special conditions already described in part II can be related mathematically to the ratio of the number of atoms reaching the oxide layer to the total number produced in the gas phase. This paper deals with the technique of the measurement of the rate of addition of hydrogen atoms to the oxide surface by a colorimetric technique. A further complication proved to be the fact that alkyl radicals formed as a result of the addition of a hydrogen atom to an olefine could be removed on the oxide layer. These radicals on removal also turned the surface blue. It is important to note that this feature of the technique permits a single radical reaction to be separated from the normal reaction sequence in the photochemical hydrogenation of an olefine. This reaction can then be studied exclusively, since the attendant complications of secondary radical reactions are entirely eliminated.

The technique further permits measurements to be made and the collision efficiencies to be computed without the necessity of measuring absolute intensities or absolute rates of production of atoms.

THEORY AND TECHNIQUE INVOLVED IN THE USE OF MOLYBDENUM TRIOXIDE AS A HYDROGEN ATOM REMOVER

In the treatment described in part II the velocity constant of the reaction of a hydrogen atom with an olefine was related mathematically with the ratio of the number of hydrogen atoms removed on the molybdenum oxide wall to the total number generated. In this treatment, no account was taken of the diminution in the concentration of the olefine due to reaction. As a result, a method had to be devised for measuring the ratio under conditions where only a small fraction of the olefine was removed.

The technique adopted permitted a measurement to be made after removing less than 0·025 mm. of olefine. The initial pressure of olefine was about 0·5 mm. and hence the reduction of concentration due to reaction was negligible. The measurement of the very small number of hydrogen atoms removed on the layer of oxide was accomplished by measuring the blueness of the surface. Qualitatively, then, since the surface becomes progressively more blue as hydrogen atoms add on, the intensity of reflected light will decrease, and this decrease can be directly related to the number of hydrogen atoms which have been removed. The apparatus employed to measure the intensity of light reflected from the surface, using a lamp of constant intensity, has been described already in part I, figure 4.

A normal run with hydrogen commences with a reflected intensity of about 15 μA, and after 15 min. the intensity drops to about 6 μA. The surface is by this time quite blue, but as seen in table 1 and figure 1, the rate of uptake of atomic hydrogen is still uniform and smooth. The surface blueness is changing but very little, and it seems that the possibility exists for the penetration of hydrogen atoms to deeper layers of the molybdenum oxide, presumably after removal on the surface layer, by some diffusion process. Thus the first point is that the surface layer alone is not entirely responsible for the removal of the hydrogen atoms and the underlying layers also play a part.

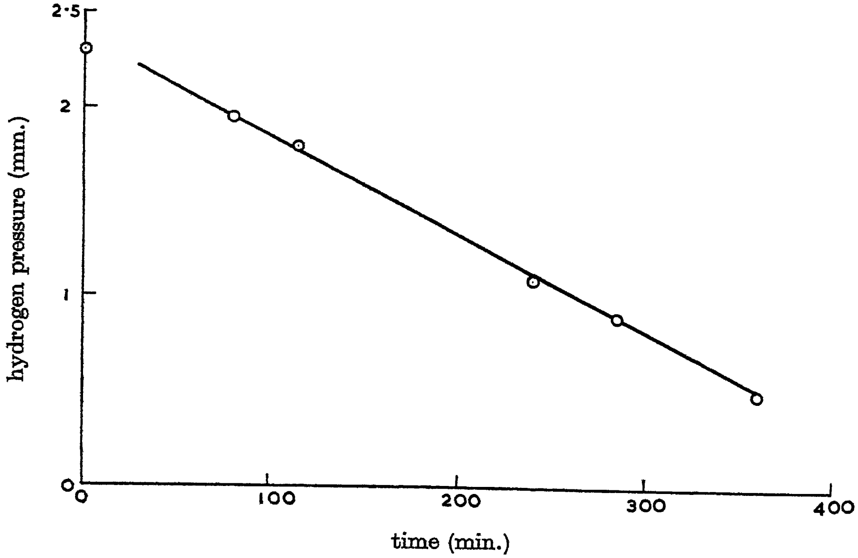


Figure 1. Uptake of hydrogen atoms by molybdenum trioxide.

The plot of a run is shown in figure 2. In this graph, the current generated by the photocell in the colorimeter is plotted against time. This curve seems to suggest some kind of reciprocal relationship between the two variables. In the following section a physical picture will be developed of what may be taking place during this period. The picture may not be entirely accurate, but it does fit the facts and provides a useful conception on which to base the deduction

Suppose the surface of the molybdenum trioxide has an area A, uniformly packed, possessing in the normal, or unreduced, state (white) a reflexion coefficient R_1. Let the intensity of incident radiation be I_0/unit area. Then the total incident radiation is

$$I = I_0 A. \tag{1}$$

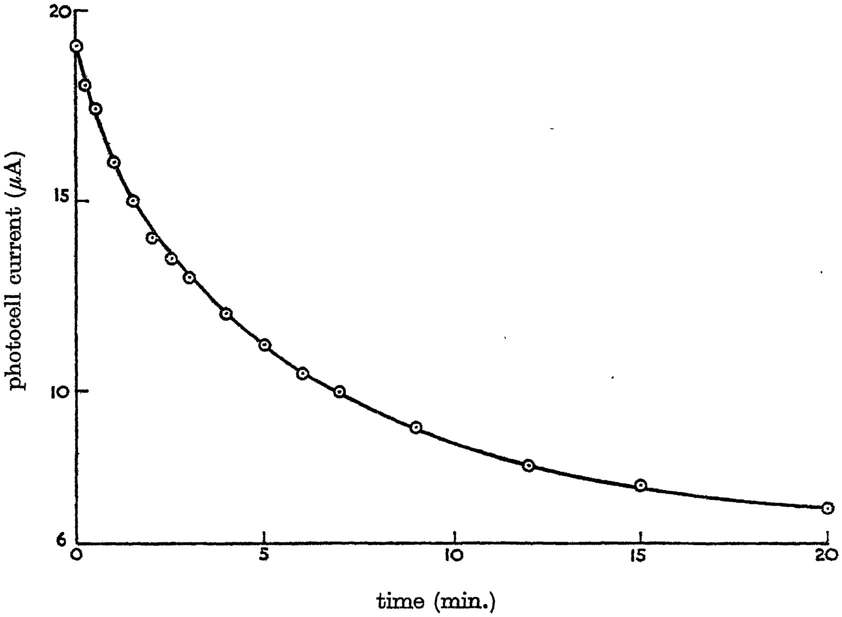


FIGURE 2. Photocell current as a function of the time of irradiation.

TABLE 1. REMOVAL OF HYDROGEN ATOMS ON MOLYBDENUM OXIDE

time (min.)	pressure of hydrogen (mm.)	photocell current (μA)
0	2·3	6·1
80	1·95	4·0
114	1·80	3·9
240	1·1	3·6
285	0·9	3·6
360	0·5	3·5

The intensity of the reflected light is then,

$$I \times R_1 = I_0 A R_1. \tag{2}$$

Now, if hydrogen atoms are generated and undergo collision with the surface, resulting in absolute removal, a new type of surface is formed, a blue surface, consisting of some unidentified reduced form of the oxide. Let the reflexion coefficient for this surface be R_2. Since the surface reflects less light,

$$R_1 > R_2.$$

As already pointed out above, the deeper layers also play a part in the removal of hydrogen atoms. The most likely mechanism is that of the hydrogen atoms removed, only those which strike a white patch will affect the blueness of the surface and may affect a deeper layer. This means that, although the rate of removal of hydrogen atoms remains constant, the rate of removal on white patches will continuously decrease.

If the atoms removed on the surface are considered to retain a certain identity, then, after a time t, the total number of hydrogen atoms removed will be Nt, where N is the number removed in unit time. Let the fraction of the surface covered by hydrogen atoms be θ, then the fraction uncovered is $(1-\theta)$. The number which can add on to the white region will be

$$Nt(1-\theta),$$

and the number migrating off a blue region will be $k\theta$, where k is a constant. Now, in the equilibrium case, the number adding on to the white surface will be equal to the number leaving the blue surface, and

$$Nt(1-\theta) = k\theta, \tag{3}$$

$$Nt = k\theta/(1-\theta). \tag{4}$$

Now the blue area is θA_0, and

$$\theta A_0 = A_0 - A_w, \tag{5}$$

where A_w is the area still unreduced to blue. Hence

$$\theta = (A_0 - A_w)/A_0, \tag{6}$$

$$1-\theta = A_w/A_0. \tag{7}$$

From (4), (6) and (7) $$Nt = k(A_0 - A_w)/A_w, \tag{8}$$

i.e. $$Kt = (A_0/A_w) - 1,$$

where $$K = N/k, \quad \text{or} \quad A_w = A_0/(1+Kt). \tag{9}$$

The intensity of the reflected light from the area A_w will be given by

$$I_w = I_0 R_1 A_w, \tag{10}$$

and from the blue area $$I_b = I_0 R_2 (A_0 - A_w). \tag{11}$$

The total intensity of the reflected radiation will be given by

$$\begin{aligned} I' &= I_w + I_b \\ &= I_0\{R_1 A_w + R_2(A_0 - A_w)\} \\ &= I_0\{A_w(R_1 - R_2) + A_0 R_2\}. \end{aligned} \tag{12}$$

Since, in the use of a photo-element with a low external resistance, the current developed is proportional to the intensity of illumination, denoting the photocell current by C',

$$C' = K'I', \tag{13}$$

where I' is the intensity of illumination and K' a constant. Substituting for I' from formula (12),

$$\begin{aligned} C' &= K'I_0\{A_0(R_1 - R_2)/(1+Kt) + A_0 R_2\} \\ &= A_0 K' I_0 (R_1 - R_2)\{1/(1+Kt) + R_2/(R_1 - R_2)\}, \end{aligned} \tag{14}$$

or $$C' = K''\{1/(1+Kt) + k''\}, \tag{15}$$

where $$K'' = A_0 K' I_0 (R_1 - R_2) \quad \text{and} \quad k'' = R_2/(R_1 - R_2).$$

Equation (15) reduces to the form

$$kt+k_1 = 1/(C'-k_2). \tag{16}$$

Since K is proportional to the number of hydrogen atoms removed on the molybdenum oxide surface in unit time, then by plotting $1/(C-k_2)$ against time, a straight line should be obtained, the slope of which will be proportional to the rate of addition of hydrogen atoms to the molybdenum oxide surface.

The technique adopted was to carry out the measurement of the value of C varying with time when only hydrogen was present in the reaction vessel from time t_1 to t_2. The reaction mixture was then altered to include a certain pressure of olefine and the run continued. Finally the run was completed by reintroducing hydrogen and continuing the measurement from t_3 to t_4. Before a linear relationship can be obtained, the value of the constant k_2 has to be determined as follows.

At time t_1,

$$kt_1+k_1 = 1/(C_1-k_2), \tag{17}$$

and at time t_2,

$$kt_2+k_1 = 1/(C_2-k_2). \tag{18}$$

Eliminating k_1 from equations (17) and (18),

$$\left.\begin{aligned} kt_1-kt_2 &= 1/(C_1-k_2)-1/(C_2-k_2), \\ k(t_1-t_2) &= (C_2-C_1)/(C_1-k_2)\,(C_2-k_2). \end{aligned}\right\} \tag{19}$$

Similarly,

$$k(t_3-t_4) = (C_4-C_3)/(C_3-k_2)\,(C_4-k_2). \tag{20}$$

Eliminating k between (19) and (20),

$$\frac{t_1-t_2}{t_3-t_4} = \frac{(C_1-C_2)\,(C_3-k_2)\,(C_4-k_2)}{(C_3-C_4)\,(C_1-k_2)\,(C_2-k_2)}. \tag{21}$$

Multiplying out equation (21) to expose terms in k_2,

$$\begin{aligned} &(t_1-t_2)\,(C_3-C_4)\,(k_2^2-(C_1+C_2)\,k_2+C_1C_2) \\ &\qquad = (t_3-t_4)\,(C_1-C_2)\,(k_2^2-(C_3-C_4)\,k_2+C_3C_4). \end{aligned} \tag{22}$$

k_2 can thus be readily determined.

The value of C can now be corrected by subtracting k_2 and a value C_c obtained. The reciprocal of this value, when plotted against time, will give a linear relationship.

As an example of this method, and to illustrate the attainable accuracy, a typical run such as that in figure 2 will be taken and corrected to obtain a linear plot. The data obtained are as set out in table 2.

To determine the value of k_2, the following points were chosen arbitrarily:

$$t_1 = 1{\cdot}0\,\text{min.}\quad C_1 = 16{\cdot}00,\qquad t_3 = 7{\cdot}0\,\text{min.}\quad C_3 = 10{\cdot}00,$$

$$t_2 = 3{\cdot}0\,\text{min.}\quad C_2 = 13{\cdot}00,\qquad t_4 = 15{\cdot}0\,\text{min.}\quad C_4 = 7{\cdot}55.$$

The value of the k_2 was determined by substituting these values in equation (22).

Figure 3 shows the corrected plot. This demonstrates the applicability of the preceding treatment. The slope of this line is the measure of the total number of hydrogen atoms being removed per unit time on the surface, and the value of $1/C_c$ is a measure of the total number removed at any time.

The next point to be cleared up is that the slope of the line obtained is proportional to the rate of addition of hydrogen atoms to the surface. This can be investigated by following the blueness of the surface for a certain time, then cutting down the intensity of radiation by interposing an intensity screen between the lamp and the

Table 2

time (min.)	photocurrent (μA)	$C_c = C - 4{\cdot}0$	$1/C_c$
0·00	19·05	15·05	0·066
0·25	18·05	14·05	0·071
0·50	17·40	13·40	0·075
1·00	16·00	12·00	0·083
1·50	15·00	11·00	0·091
2·00	14·05	10·05	0·099
2·50	13·50	9·50	0·105
3·00	13·00	9·00	0·111
4·00	12·05	8·05	0·124
5·00	11·25	7·25	0·138
6·00	10·50	6·50	0·154
7·00	10·00	6·00	0·167
9·00	9·05	5·05	0·198
12·00	8·05	4·05	0·247
15·00	7·55	3·55	0·282
20·00	6·95	2·95	0·339

The value of k_2 used to obtain the value of C_c was 4·00.

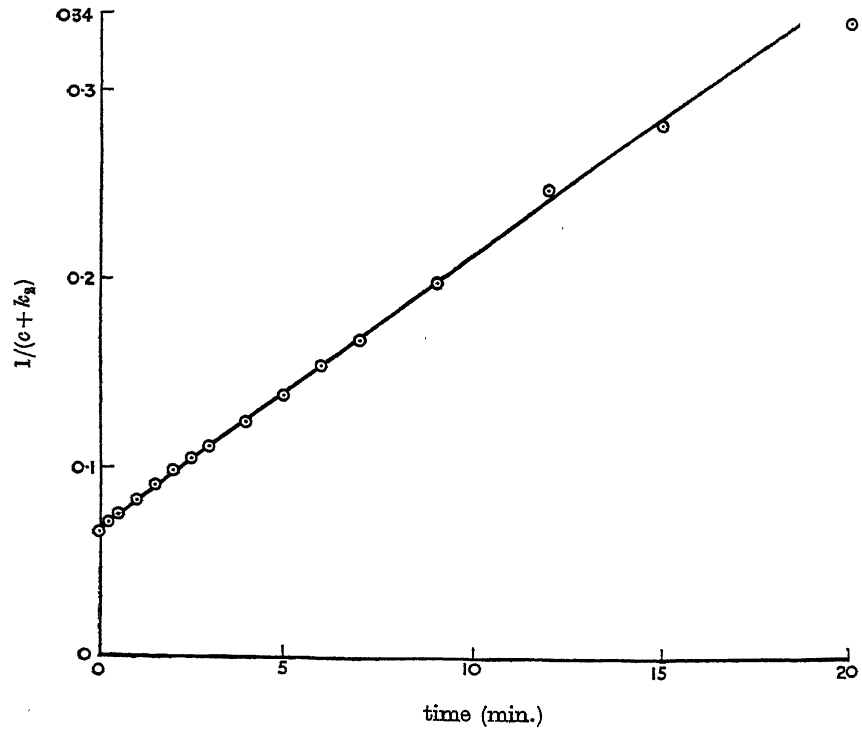


Figure 3. Corrected curve of figure 2.

reaction vessel. By cutting down the intensity by a factor R', it would be expected that the slope of the lines obtained in the two cases would differ by a factor R'.

Run 240, the data for which are shown in table 3, was conducted for the first 3 min. without an intensity screen; the screen was then inserted and the run continued for a further 7 min. The value of k_2 was found to be 5·4. The ratio of the two slopes was 0·39 and the transmission factor of the screen about 0·35. This agreement is considered to be satisfactory. The plot is shown in figure 4.

TABLE 3. EFFECT OF INTENSITY OF IRRADIATION ON RATE OF BLUEING

time (min.)	C	C_c	$1/C_c$
0	17·55	12·15	0·082
0·5	13·9	8·15	0·118
1·0	12·0	6·6	0·152
1·5	10·9	5·5	0·182
2·0	9·9	4·5	0·222
2·5	9·45	4·05	0·247
3·0	8·95	3·55	0·282
4·0	8·60	3·20	0·313*
5·0	8·35	2·95	0·339
6·0	8·15	2·75	0·364
7·0	7·95	2·55	0·392
8·0	7·65	2·25	0·444
9·0	7·5	2·1	0·476

* Intensity screen inserted after 3 min.

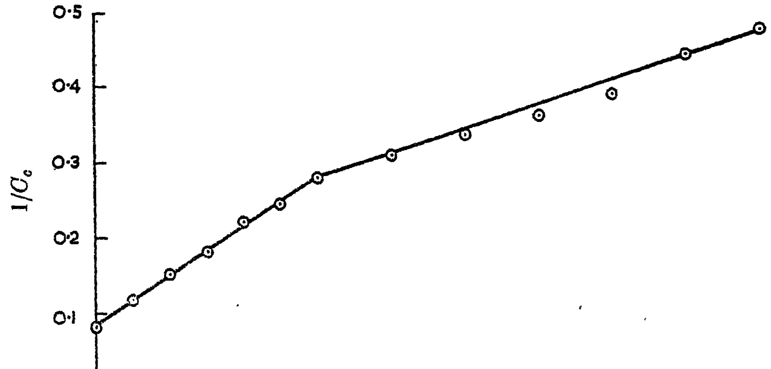


FIGURE 4. Run with intensity screen.

It is then clear that the slope of the curve plotted from the values of $1/C_c$ against t provides a measure of the number of hydrogen atoms being removed on the oxide surface in unit time.

It is a well-known fact (Neujmin 1937; Taylor & Bates 1930) that radicals other than hydrogen atoms can add on to a surface of molybdenum oxide or tungsten oxide, so that the possibility had to be considered that the alkyl radicals produced by the primary addition of a hydrogen atom to an olefine might add on to the oxide surface as well and affect the blueness of the surface. By decomposing mercury dibutyl in the cylindrical reaction vessel by 2537 Å radiation, the rate of increase of pressure is a measure of the rate of decomposition of the dialkyl. The results of such runs are shown in table 4 and figure 5. Runs 175 and 176 are then a measure of the rate of decomposition of the alkyl which was present in the reaction vessel at its vapour pressure, liquid dialkyl being present throughout the experiments in excess to maintain the vapour pressure. The reason for the decrease in rate with time is possibly due to deactivation of the excited mercury atoms by the increasing amount of hydrocarbon products of decomposition, but this need not affect the general conclusions drawn here. Runs 177 and 178 were done in the same reaction vessel under the same conditions, except that in this case, the bottom of the reaction vessel was covered with a thin layer of powdered molybdenum oxide. The pressure in this case rises much more slowly. At the same time, the oxide turns blue, indicating that something is being removed, and that the effect on the surface blueness is not to be neglected.

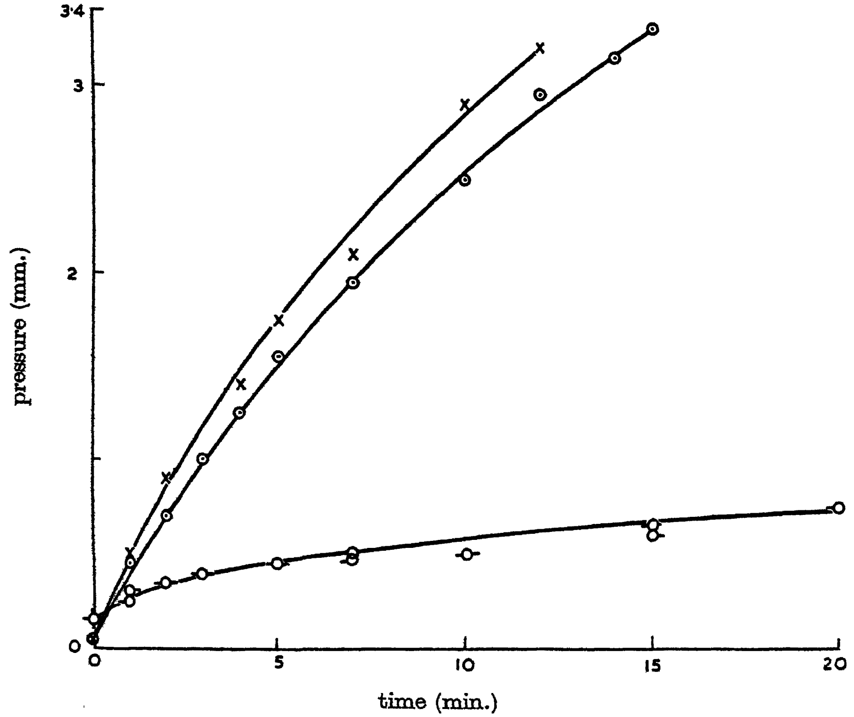


Figure 5. Decomposition of mercury dibutyl. × Run 175; ⊙ run 176; ○- run 177; -○ run 178.

Under these experimental conditions, the alkyl radicals had to diffuse some 20 mm. before colliding with the oxide surface, but nevertheless, about 85 % of those formed retained their identity and ability to react on the surface of the oxide and be removed from the gas phase. Under these conditions, which will be defined later, this path length is reduced to less than 2·5 mm. It is reasonable to infer that when the lifetime in the gas phase is so reduced, all the radicals produced will be removed by interaction on the molybdenum oxide surface. Further proof of the correctness of this will be demonstrated later.

TABLE 4. DECOMPOSITION OF MERCURY DIBUTYL WITH 2537 Å RADIATION

time (min.)	clean reaction vessel, pressure (mm.)		reaction vessel with layer of MoO_3, pressure (mm.)	
	run 175	run 176	run 177	run 178
0	0·06	0·05	0·15	0·15
1	0·50	0·45	0·30	0·25
2	0·90	0·70	0·35	0·35
3	1·15	1·0	0·40	0·40
4	1·40	1·25	—	—
5	1·75	1·55	0·45	0·45
7	2·10	1·95	0·50	0·48
10	2·90	2·50	0·50	0·50
12	3·20	2·95	—	—
14	3·65	3·15	—	—
15	3·80	3·30	0·60	0·65

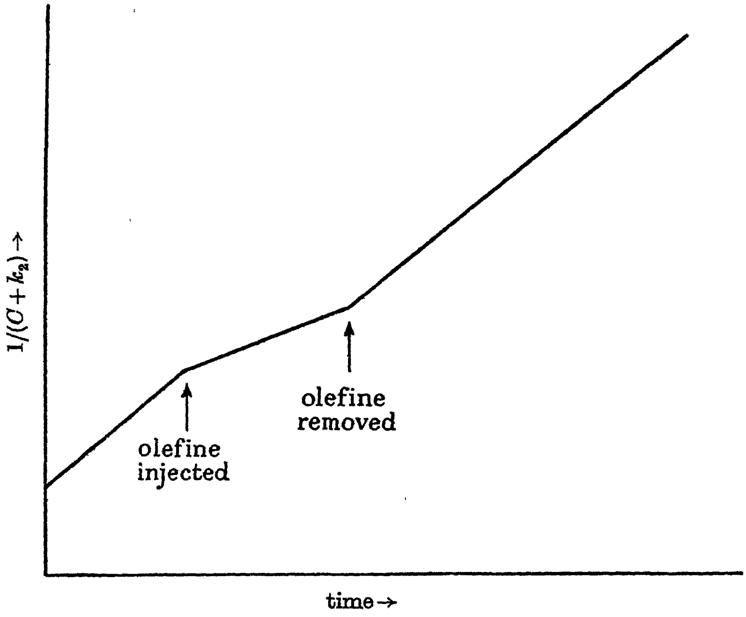


FIGURE 6. Type of experimental run to be expected.

Suppose a run is conducted for a certain time with hydrogen alone, then with an olefine present, and finally again with hydrogen. If the values of the current generated by the photocell are corrected by the determination and use of the constant k_2, then a graph should be obtained as shown in figure 6. The ratio of the slopes of the different parts of the curve gives the ratio of the number of atoms reaching the surface of oxide to the total number of atoms produced. This presupposes that the alkyl radicals removed on the oxide surface have no significant effect on the blueness of the surface. However, it is known that this is not necessarily the case and allowance must be made.

Let the blueing effect of a hydrogen atom be k_{H} and the effect of an alkyl radical be $k_{\mathrm{H}}.k_R$. Then if N' atoms are generated in unit time, the blueing rate will be given by

$$R_{\mathrm{H}} = k_{\mathrm{H}}N'. \tag{23}$$

If the fraction reaching the removing surface is N'_{H}/N', then the rate of blueing due to hydrogen atoms will be

$$R'_{\mathrm{H}} = k_{\mathrm{H}}N'_{\mathrm{H}} \tag{24}$$

and the rate due to the alkyl radicals, one of which will be formed for each hydrogen atom removed in the gas phase, is given by

$$R'_R = k_R k_{\mathrm{H}}(N' - N'_{\mathrm{H}}). \tag{25}$$

The observable rate will then be given by

$$\begin{aligned} R' &= R'_{\mathrm{H}} + R'_R \\ &= k_{\mathrm{H}}N'_{\mathrm{H}}\{1 + k_R(N'/N'_{\mathrm{H}} - 1)\}. \end{aligned} \tag{26}$$

The rate for hydrogen alone is given by equation (23), and the ratio of the two rates, A', a measurable quantity will then be expressed by

$$\left.\begin{aligned} A' &= R'/R_{\mathrm{H}} \\ &= N'_{\mathrm{H}}\{1 + k_R(N'/N'_{\mathrm{H}} - 1)\}/N', \\ A' &= N'_{\mathrm{H}}(1 - k_R)/N' + k_R, \\ N'_{\mathrm{H}}/N' &= (A' - k_R)/(1 - k_R). \end{aligned}\right\} \tag{27}$$

Now, N'_{H}/N' is the fraction of the atoms produced which get to the removing surface. Suppose the experiment is carried out again using a different path distance, i.e. placing the removing surface at a different distance from the silica window, then the following will be the expression obtained:

$$N''_{\mathrm{H}}/N'' = (A'' - k_R)/(1 - k_R). \tag{28}$$

By taking the path distance to be (*a*) 1·25 mm. and (*b*) 2·5 mm. respectively, it can be seen from part II, figure 4, that the ratio of N'_{H}/N' to N''_{H}/N', i.e. R' to R'', varies with the value of b from 1 to approximately 2 and that for the value of b in a particular reaction, there exists also a singular and characteristic value for the ratio R'/R'',

$$R'/R'' = (A' - k_R)/(A'' - k_R). \tag{29}$$

It is, of course, necessary that the value of D remains constant in both experiments, but by using the same gas pressures in both, this is readily achieved.

If a random value is taken for R'/R'' between 1 and 2, the corresponding value for k_R can be obtained from equation (29) since A' and A'' are determined experimentally. This value can then be substituted in equation (27) and the value of R' obtained. From the curve in part II, figure 4, the corresponding value of b is obtained. The value of R'' can then be obtained from the curve for the appropriate value of b.

The new ratio of R'/R'' which can now be obtained will form a better approximation to the true value and a better value for k_R can be obtained on substitution in equation 29. By progressing in this way, a number of successively closer approximations can be obtained which converge on the value satisfying the conditions.

The convergence can be obtained either by repeated and exhaustive use of the above method, or by obtaining several values for k_R as above and evaluating the corresponding values of R' and R'', thence obtaining the values of b' and b''. These can be plotted against the values of k_R used and the point of intersection of the two curves gives the true value of b and k_R. The graphical method is rather inaccurate as the relationship of b to k_R is not linear. However, a mechanical method will be described below which greatly simplifies the labour of this type of calculation.

Perhaps an example will serve to clarify the method just described in the past few pages.

Suppose for a mixture of constant composition, and constant diffusion coefficient D that the value of b is 8·0, then referring to part II, figure 4, the corresponding ratio R' at a separation of 1·25 mm. is 0·745, which means that the number of hydrogen atoms reaching the layer is $0{\cdot}745N$ and the number of radicals is $0{\cdot}255N$.

The rate of blueing will be given by R',

$$\begin{aligned} R' &= k_H \times 0{\cdot}745N + k_R k_H \times 0{\cdot}255N \\ &= k_H N(0{\cdot}745 + 0{\cdot}255 \times k_R). \end{aligned}$$

The corresponding rate for hydrogen alone would be

$$R_H = k_H N.$$

Then $$A' = 0{\cdot}745 + 0{\cdot}255 k_R.$$

Let the value of k_R be 1·2, i.e. the radical is 1·2 times as effective as a hydrogen atom in turning the surface blue.

This gives a value for A' of 1·051.

In the same way, for a separation of 2·5 mm., the corresponding value of N''_H/N'' for $b = 8{\cdot}0$ is 0·42 which leads to a value of

$$A'' \doteq 1{\cdot}115.$$

By carrying out a run, the only observable quantities would be

$$A' = 1{\cdot}051, \text{ for spacing of } 1{\cdot}25\,\text{mm.}$$
$$A'' = 1{\cdot}115, \text{ for spacing of } 2{\cdot}50\,\text{mm.}$$

By the method of successive approximation, applying equation (29),

$$(N'_H/N')/(N''_H/N'') = (A' - k_R)/(A'' - k_R).$$

Now, take an average value for the ratio of the left-hand side of the equation. Take 1·5 as already suggested above, then

$$1{\cdot}5 = (1{\cdot}051 - k_R)/(1{\cdot}115 - k_R),$$

from which, $$k_R = 1{\cdot}242.$$

Substitute for k_R in equation (27),

$$N'_{\mathrm{H}}/N' = 0{\cdot}79.$$

For this value the corresponding value for N''_{H}/N'' would be 0·48 (see part II, figure 4) so that the new approximation for $(N'_{\mathrm{H}}/N')/(N''_{\mathrm{H}}/N'')$ is 0·79/0·48 which is 1·65.

Applying this new and more accurate value to equation 29, the new value of k_R is obtained,

$$k_R = 1{\cdot}21.$$

Substituting for k_R in (27), $N'_{\mathrm{H}}/N' = 0{\cdot}76.$

The corresponding value for N''_{H}/N'' is 0·435 which gives the value for

$$(N'_{\mathrm{H}}/N')/(N''_{\mathrm{H}}/N'') = 1{\cdot}75.$$

Again substituting this value for N'_{H}/N' in equation (29), the value for k_R is 1·2, and by use of equation (27), the value of N'_{H}/N' is found to be 0·75.

This gives a value for N''_{H}/N'' of 0·425; therefore the new value of $(N'_{\mathrm{H}}/N')/(N''_{\mathrm{H}}/N'')$ will be 1·76. This is so close to the last value that the solution has been obtained. Thus

$$k_R = 1{\cdot}2, \quad N'_{\mathrm{H}}/N' = 0{\cdot}75,\ b = 8{\cdot}0, \quad N''_{\mathrm{H}}/N'' = 0{\cdot}425,\ b = 7{\cdot}9.$$

The graphical method is as follows.

After making the first approximations of $(N'_{\mathrm{H}}/N')/(N''_{\mathrm{H}}/N'') = 1{\cdot}5$ and obtaining $k_R = 1{\cdot}24$, substituting in equations (27) and (28), the values of N'_{H}/N' and N''_{H}/N'' can be obtained.

$$N'_{\mathrm{H}}/N' = 0{\cdot}79, \qquad N''_{\mathrm{H}}/N'' = 0{\cdot}525,$$
$$b = 7{\cdot}0, \qquad b = 6{\cdot}3.$$

After the second approximation, $k_R = 1{\cdot}21$,

$$N'_{\mathrm{H}}/N' = 0{\cdot}76, \qquad N''_{\mathrm{H}}/N'' = 0{\cdot}45,$$
$$b = 7{\cdot}7, \qquad b = 7{\cdot}4.$$

After the third approximation, $k_R = 1{\cdot}20$,

$$N'_{\mathrm{H}}/N' = 0{\cdot}75, \qquad N''_{\mathrm{H}}/N'' = 0{\cdot}425,$$
$$b = 8{\cdot}0 \qquad b = 7{\cdot}9.$$

These results are tabulated in table 5. By plotting these results as in figure 7 the point of intersection can be determined and the value of b obtained.

TABLE 5

k_R	b'	b''
1·24	7·0	6·3
1·21	7·7	7·4
1·20	8·0	7·9

This method is not always applicable since the plot of b against k_R need not be linear and the extrapolation of the curves is not a very simple matter. Further, in the example shown above, the values have been chosen for simplicity of manipulation and, in practice, it is rare that only three approximations have to be used. The method can then be extremely laborious and time-consuming. A mechanical method was devised to overcome these mathematical difficulties and it will now be described.

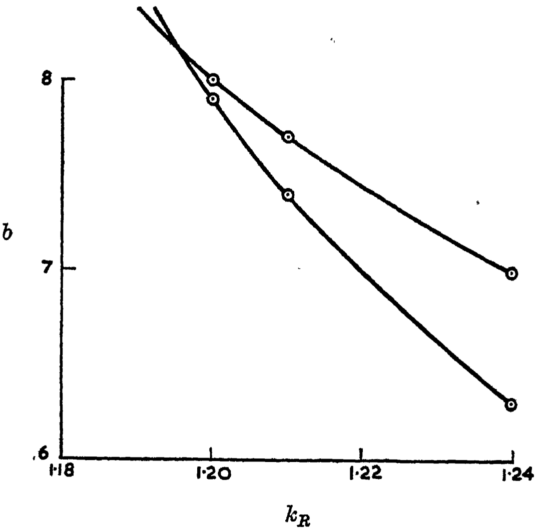


FIGURE 7. Graphical method of solution for b.

MATHEMATICAL CALCULATOR

Theory

When account is taken of the effect of alkyl radicals in the blueing of molybdenum oxide, the equations used are

$$N'_{\text{H}}/N' = (A'-k_R)/(1-k_R), \tag{27}$$

and

$$N''_{\text{H}}/N'' = (A'-k_R)/(1-k_R), \tag{28}$$

where the values of A' and A'' represent the experimentally observed blueness ratios mentioned above and k_R is a constant depending on the radical generated by the addition of a hydrogen atom to an olefine. Further, the value is such that N'_{H}/N' and N''_{H}/N'' when referred to their respective curves in part II, figure 4, will both give the same value for b. Thus the problem is to alter k_R until these conditions are satisfied.

In figure 8, let the length AC be 1 unit. CE is perpendicular to AD. Let AH/AC represent the value of A' and AK/AC the value of A''. Erect the lines KK' and HH' perpendicular to AD. Now, let the line AB move along AC and CE so that its length remains constant and let AF/FC be the value of k_R. Then

$$A'-k_R = FH/AC, \quad A''-k_R = FK/AC, \quad 1-k_R = FC/AC$$

$$N'_{\text{H}}/N' = (A'-k_R)/(1-k_R) = FH/AC.AC/FC = FH/FC,$$

also

$$N''_{\text{H}}/N'' = FK/FC.$$

By geometry of the system,

$$FH/FC = FM'/FG, \quad FK/FC = FN'/FG,$$

so that if FG is unity on a new scale,

$$N'_{\mathrm{H}}/N' = FM' \quad \text{and} \quad N''_{\mathrm{H}}/N'' = FN'.$$

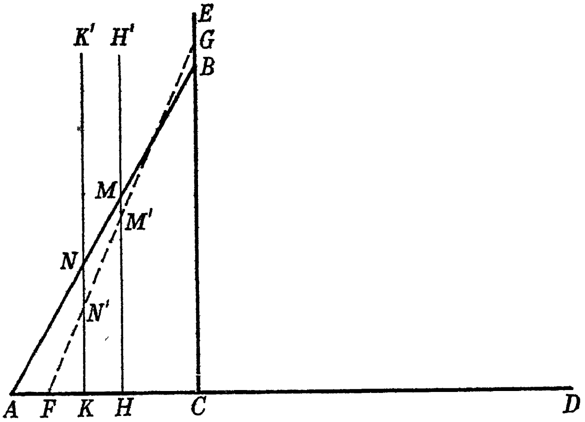


FIGURE 8. Theoretical diagram of calculator.

By affixing to the sliding member FG, a calibration scale relating the fraction of the length of the member from F with the corresponding value of b for the two values of the path distance, there are then two scales on the sliding arm. By setting the experimentally observed values of A' and A'' on the sliding cursors KK' and HH' and moving the arm AB until the reading indicated by the intersection of the two cursors on their respective scales is the same, this value is the required value of b and the distance AF measured on another scale is the value of k_R.

Construction of the machine

The apparatus is illustrated in figure 9.

The two cursors move on two rods A, of $\frac{1}{4}$ in. duralumin rod. The inset I shows the details of the travelling ends of the cursors. These are made from brass tube and sheet, the small pointers of which have a small hole and slot as indicated to facilitate the fitting and removal of fresh cursor threads. The cursors are thin black linen threads, maintained taut by two light springs C of half-hard brass wire. The two cursor scales B must be lined up accurately parallel and when the thread is reading the same on both scales, the thread must be parallel to the vertical tube and at right angles to the scales. Twin scales are used to provide a simple means for setting the cursors to a particular value and keeping the thread parallel to the vertical tube. The scales are mounted on wooden dowel-rods and are slotted so that final adjustments can be made.

The runners E and F are of $\frac{1}{2}$ in. brass tube fastened down with hoops of brass. The sliding parts H are constructed by fitting brass tubing snugly to the size of E by inserting liners of thin brass until the desired fit is obtained. The lower sliding end

H_2 carries a pointer lined up to the point about which the bar G pivots. The bar G is made accurately 20 cm. between the centres of the pivots on H_1 and H_2 and is held in place by two countersunk screws fitted to hold the bar securely, but to allow easy rotation about the screw.

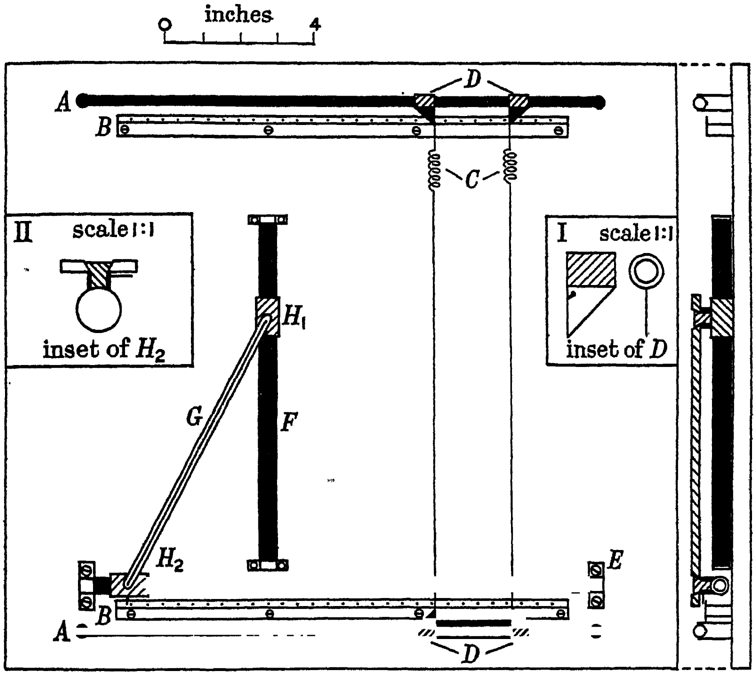


FIGURE 9. Construction of calculator.

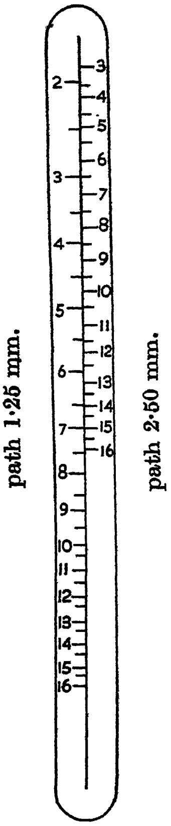


FIGURE 10. Scale fixed to calculator.

F is lined up perpendicular to A and its midline opposite the 10 cm. division of scales BB. E is run parallel to A. The scales on G are constructed from figure 7 and are divided by marking the values of b at the point on the scale such that FM'/FG is the required ratio determined from the graph. The scales are seen in figure 10. Now, by measuring the values of the ratios A' and A'' and setting them on the cursors, with the pointer on H_2 at zero, the runner H_2 can be slid along until the readings on the scale on G indicated by the respective cursors are the same. This is then the value of b required and the factor k_R is indicated on the lower scale by the pointer on H_2.

This instrument reduces the time required for a calculation of this kind from perhaps 40 min. to 30 sec.

REFERENCES

Neujmin, H. 1937 *C.R. Acad. Sci. U.R.S.S.* **16**, 447.
Taylor, A. S. & Bates, J. R. 1930 *Nature*, **125**, 599.

The kinetics of the interaction of atomic hydrogen with olefines

IV. Results obtained by the foregoing techniques

By H. W. Melville, F.R.S. and J. C. Robb
Chemistry Department, Marischal College, Aberdeen

(*Received* 23 *June* 1948—*Read* 10 *February* 1949)

The experimental techniques and theoretical considerations developed in the previous parts have been utilized in obtaining data on the reactions of the type $H + C_nH_{2n} = C_nH_{2n+1}$. The olefines used were ethylene, propylene, *iso*-butene, *n*-pentene-2 (*cis*), and 2, 3, 3, tri-methyl-butene-1. Certain tentative suggestions as to the effect of structure on the reactivity of the double bond are put forward, but in general, the collision efficiency of a hydrogen atom with an olefine lies between 10^{-3} and 10^{-4}.

It is noted that in the computation of the collision efficiency, it seems to be rather unsatisfactory that the diameter of the whole molecule should be used and it is suggested that perhaps the effective diameter of a more intimate part of the molecule in the region of the double bond should be employed.

INTRODUCTION

In the previous parts, the apparatus has been described and the method of differential equations applied to the solution of the necessary diffusion theory. The colorimeter technique for the estimation of the rate of addition of hydrogen atoms to the oxide surface has been developed and the effect of alkyl radicals on the oxide blueness allowed for.

In this paper, some results obtained by this technique are given. These results, although limited in number, permit a certain tentative correlation of the data obtained. The olefines are discussed in the following order: ethylene, propylene, *iso*-butene, *n*-pentene-2 (*cis*), 2, 3, 3, tri-methyl-butene-1; and observations peculiar to a specific olefine are included in the relevant sections.

ETHYLENE

It should be noted here that the technique involves carrying on a blueing run for 2·5 min. with hydrogen alone, then introducing a small pressure of olefine by opening the injector and flowing the mercury up and down several times. This ensures complete mixing of olefine and hydrogen. The run is then continued for a further 5·5 min., and the gas mixture pumped out completely. The appropriate pressure of hydrogen is then introduced and the blueing continued for a further 7 min. From the values obtained during the two 'hydrogen' periods, the value of the corrective constant k_2 is obtained. All the observed values of photo-cell current are then corrected by the subtraction of k_2 and the reciprocal of these values plotted against time. The ratio of the first two portions of the graph is then used to calculate directly the collision efficiency with the aid of the calculating machine already described in part III.

Since at $t = 2·5$ and $t = 8$ min., a physical manipulation of the gases occurs, it can occur that the oxide surface becomes slightly altered as a result of the gas flow. This has the effect of shifting the next portion of the graph bodily either up or down, but does not alter the resulting slope. It is not then surprising that the curves are not necessarily continuous at the times mentioned above. The collision efficiency of the reaction

$$H + C_2H_4 = C_2H_5$$

was measured by carrying out the run at two spacings, 1·25 and 2·5 mm., using tungsten oxide as a removing surface. Table 1 gives the data obtained in runs 335 and 336 for a spacing of 2·5 mm.

TABLE 1. RUNS WITH ETHYLENE—SPACING 2·5 MM.

time (min.)	run 335 C	run 335 C_c	run 335 $1/C_c$	run 336 C	run 336 C_c	run 336 $1/C_c$	
0·0	15·0	10·6	0·094	16·0	11·7	0·085	A
0·5	14·0	9·6	0·104	14·95	10·65	0·094	
1·0	13·4	9·0	0·111	13·9	9·6	0·104	
1·5	12·7	8·3	0·120	13·4	9·1	0·110	
2·0	12·3	7·9	0·127	12·9	8·6	0·116	
2·5	11·9	7·5	0·133	12·4	8·1	0·123	
2·5	11·9	7·5	0·133	12·25	7·95	0·126	B
3·0	11·6	7·2	0·139	12·0	7·7	0·130	
4·0	11·3	6·9	0·145	11·5	7·2	0·139	
5·0	10·95	6·55	0·153	11·0	6·7	0·149	
6·0	10·5	6·1	0·164	10·55	6·25	0·160	
7·0	10·40	6·00	0·167	10·4	6·1	0·164	
8·0	10·05	5·65	0·177	10·0	5·7	0·175	
8·0	9·85	5·45	0·183	10·0	5·7	0·175	C
9·0	9·5	5·1	0·196	9·6	5·3	0·189	
10·0	9·05	4·65	0·215	9·4	5·1	0·196	
12·0	8·5	4·1	0·244	8·6	4·3	0·233	
15·0	7·9	3·5	0·286	7·9	3·6	0·278	

Run 335: Time A: 6·2 mm. hydrogen in reaction vessel.
Time B: 0·5 mm. ethylene injected.
Time C: 6·0 mm. hydrogen in reaction vessel.

Run 336: Time A: 6·1 mm. hydrogen in reaction vessel.
Time B: 0·5 mm. ethylene injected.
Time C: 6·0 mm. hydrogen in reaction vessel.

C is the photocell current in μA, C_c is the value corrected by k_2 and $1/C_c$ the reciprocal of the corrected value. The correction factor k_2 is calculated by using the formula (22) derived in part III. The values substituted in the equation were those at $t_1 = 0·5$, $t_2 = 2·5$, $t_3 = 9·0$ and $t_4 = 15·0$ min.

The value of k_2 for run 335 is 4·4, and for run 336 the value is 4·3. These two runs are plotted in figure 1. For clarity, run 335 is displaced 0·01 unit upwards. For run 335, the slope of the first part is taken between $t = 0·5$ and $t = 2·5$ min. and is

$$(0·133 - 0·104)/2 = 0·0145.$$

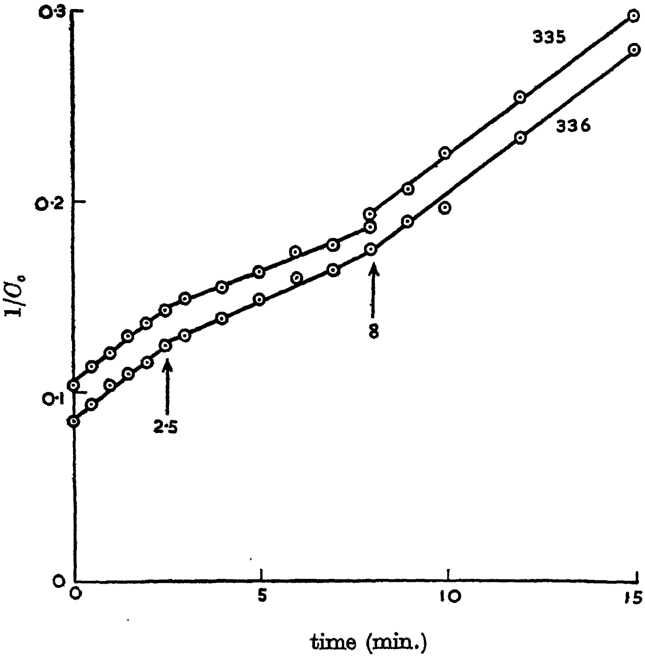


FIGURE 1. Runs with ethylene—spacing 2·5 mm. (Tungsten oxide).

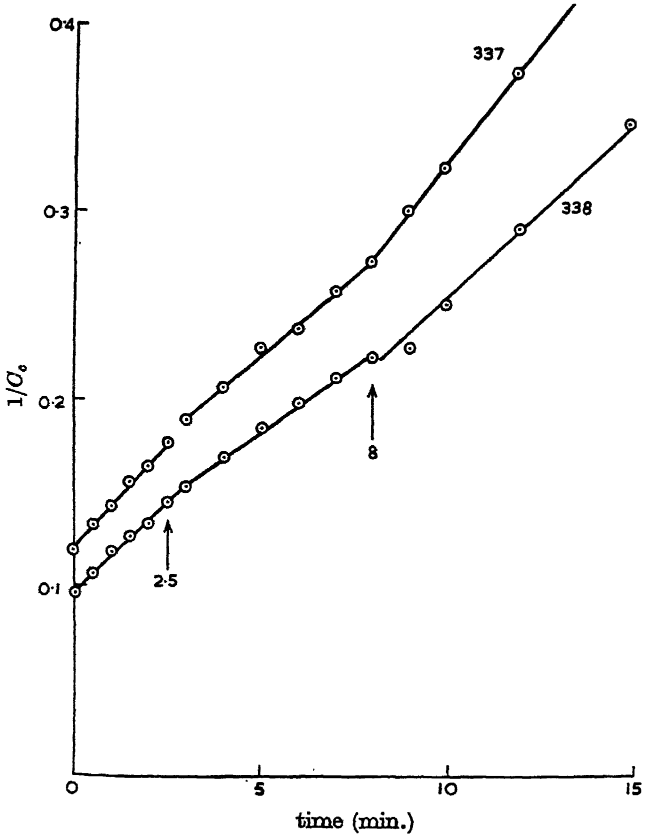


FIGURE 2. Runs with ethylene—spacing 1·25 mm. (Tungsten oxide).

The slope for the second part is taken between $t = 3{\cdot}0$ and $t = 8{\cdot}0$ min. The value of the slope is

$$(0{\cdot}177 - 0{\cdot}139)/5 = 0{\cdot}0076.$$

The ratio of these two slopes is 0·53.

In a similar manner, the ratio of the two slopes in run 336 is 0·56. This gives an average slope of 0·545.

Table 2 gives the data obtained for runs 337 and 338 carried out at a spacing of 1·25 mm. using again tungsten oxide as a removing surface. From the data in this table, for run 337 the value of k_2 is 7·0 and for run 338, the value of k_2 is 6·1. The plots of the corrected curves are given in figure 2. In this figure, run 337 is displaced 0·01 upwards. The ratios of the two slopes in the two runs are respectively 0·76 and 0·72, giving an average ratio of 0·74.

Thus for a spacing of 2·5 mm., the ratio is 0·54 and for a spacing of 1·25 mm., the ratio is 0·74.

TABLE 2. RUNS WITH ETHYLENE—SPACING 1·25 MM.

time	run 337			run 338			
(min.)	C	C_e	$1/C_e$	C	C_e	$1/C_e$	
0·0	16·4	9·4	0·106	16·45	10·35	0·097	A
0·5	15·1	8·1	0·123	15·45	9·35	0·107	
1·0	14·5	7·5	0·133	14·5	8·4	0·119	
1·5	13·85	6·85	0·146	14·0	7·9	0·127	
2·0	13·5	6·5	0·154	13·55	7·45	0·134	
2·5	13·0	6·0	0·167	13·0	6·9	0·145	
2·5	13·0	6·0	0·167	13·0	6·9	0·145	B
3·0	12·6	5·6	0·179	12·6	6·5	0·154	
4·0	12·1	5·1	0·196	12·0	5·9	0·169	
5·0	11·6	4·6	0·217	11·5	5·4	0·185	
6·0	11·4	4·4	0·227	11·15	5·05	0·198	
7·0	11·05	4·05	0·247	10·85	4·75	0·211	
8·0	10·8	3·8	0·263	10·6	4·5	0·222	
8·0	10·8	3·8	0·263	10·6	4·5	0·222	C
9·0	10·45	3·45	0·290	10·5	4·4	0·227	
10·0	10·2	3·2	0·313	10·1	4·0	0·250	
12·0	9·75	2·75	0·364	9·55	3·45	0·290	
15·0	9·1	2·1	0·476	9·0	2·9	0·345	

Run 337: Time A: 6·0 mm. hydrogen in reaction vessel.
Time B: 0·5 mm. ethylene injected.
Time C: 5·8 mm. hydrogen in reaction vessel.

Run 338: Time A: 5·9 mm. hydrogen in reaction vessel.
Time B: 0·5 mm. ethylene injected.
Time C: 6·0 mm. hydrogen in reaction vessel.

By setting these two values on the calculator and sliding the scale until the same value of b is obtained on both scales when read against their respective cursors, the value of b is found to be 12 and the corresponding value for k_R is 0·38.

Since, by definition,

$$b = (g/D)^{\frac{1}{2}}, \quad g = 144D. \tag{1}$$

D is obtained as shown in part II of this series and is calculated to be $1\cdot 14\times 10^4$ cm.^{2}/sec., hence

$$g = 1\cdot 65\times 10^4.$$

Further, by definition,

$$g = kn, \tag{2}$$

where k is the velocity constant and n the number of olefine molecules per ml. In 1 ml. of gas at a pressure of 1 mm. there are $3\cdot 6\times 10^{16}$ molecules. Then substituting in equation (2)

$$\begin{aligned} k &= 1\cdot 65\times 10^4/(0\cdot 5\times 3\cdot 6\times 10^{16}) \\ &= 9\cdot 1\times 10^{-13}\,\text{ml.mol.}^{-1}\text{sec.}^{-1}. \end{aligned}$$

The collision efficiency is obtained by dividing the velocity constant, k, by the collision number. The collision number is given by

$$Z = \sigma^2\{8RT(M_{\mathrm{H}}+M_{\mathrm{X}})/M_{\mathrm{H}}\,.\,M_{\mathrm{X}}\}^{\frac{1}{2}},$$

where σ is the collision diameter of the two species involved, R is the gas constant, T the absolute temperature, M_{H} the mass of the hydrogen atom and M_{X} the mass of the olefine molecule. The values of Z for the various olefines used is given in table 3.

TABLE 3. COLLISION NUMBER FOR H + OLEFINE

olefine	collision number
ethylene	$1\cdot 1\times 10^{-9}$
propylene	$1\cdot 6\times 10^{-9}$
i-butene	$1\cdot 5\times 10^{-9}$
n-pentene-2 (*cis*)	$2\cdot 1\times 10^{-9}$
2, 3, 3, tri-methyl-butene-1	$1\cdot 9\times 10^{-9}$

TABLE 4. RUN WITH ETHYLENE—SPACING 4·0 MM.

time	run 321			
(min.)	C	C_c	$1/C_c$	
				A
0·0	14·9	12·3	0·081	
0·5	13·9	11·3	0·088	
1·0	13·0	10·4	0·096	
1·5	12·5	9·9	0·101	
2·0	12·0	9·4	0·106	
2·5	11·45	8·85	0·113	
				B
2·5	11·5	8·9	0·112	
3·0	11·2	8·6	0·116	
4·0	10·9	8·3	0·120	
5·0	10·55	7·95	0·126	
6·0	10·4	7·8	0·128	
7·0	10·0	7·4	0·135	
				C
7·0	10·1	7·5	0·133	
8·0	9·55	6·95	0·144	
10·0	8·5	5·9	0·169	
12·0	7·8	5·2	0·192	
15·0	6·95	4·35	0·230	

Time A: 6·2 mm. hydrogen in reaction vessel.
Time B: 0·5 mm. ethylene injected.
Time C: 8·0 mm. hydrogen in reaction vessel.

Hence, for ethylene,

$$\text{collision efficiency} = 9 \cdot 1 \times 10^{-13}/1 \cdot 1 \times 10^{-9}$$
$$= 8 \cdot 3 \times 10^{-4}.$$

Referring to part II, figure 4, it is seen that if a path-distance of 4 mm. is employed, virtually all the hydrogen atoms would be removed before reaching the oxide layer. Certainly, at the most, only a fraction of about 0·05 could get to the layer. This would mean that 0·95 of the active centres removed would be ethyl radicals, assuming that no radical recombination occurred in the gas-phase. Then the blueing rate in presence of ethylene would be

$$0 \cdot 95 \times k_R + 0 \cdot 05,$$

referred to the hydrogen rate as unity. Since the radical blueing coefficient has been shown to be 0·38, this gives a relative rate of 0·41. Thus, the ratio of the two parts of the run done at this spacing should lie between 0·41 and 0·38.

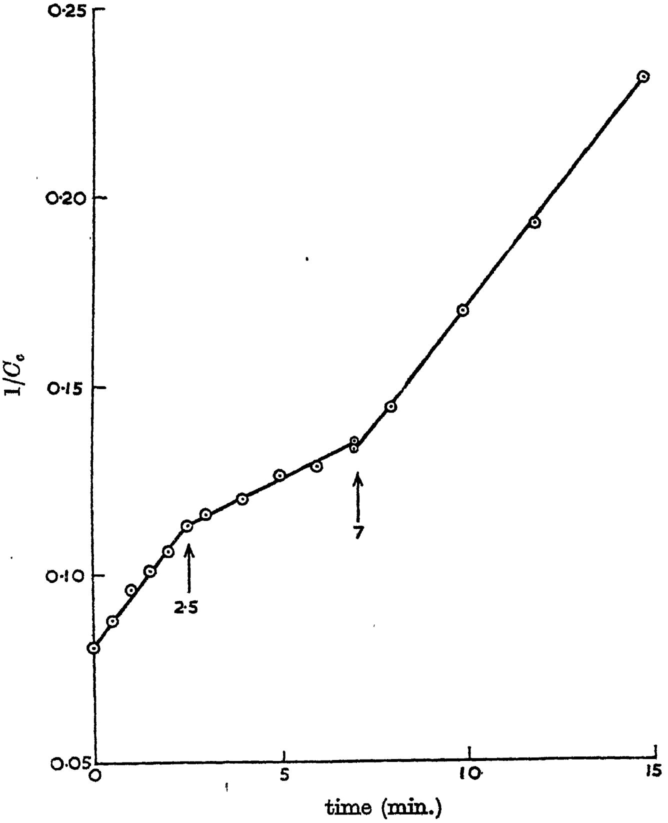


FIGURE 3. Run with ethylene—spacing 4·0 mm. (Tungsten oxide).

In table 4 are given the data obtained from run 321, at a spacing of 4 mm.

The corrective factor is found to be 2·6. The plot of the corrected curve is shown in figure 3. The ratio of the two slopes is obtained from the values at $t = 0$, $t = 2 \cdot 5$, $t = 2 \cdot 5$ and $t = 7 \cdot 0$ min. The slope of the first part is given by

$$(0 \cdot 113 - 0 \cdot 081)/2 \cdot 5 = 0 \cdot 0128,$$

and the slope of the second part is given by

$$(0{\cdot}135-0{\cdot}113)/4{\cdot}5 = 0{\cdot}0049.$$

The ratio of these two slopes is then 0·38.

This is good evidence for the fact that there is no recombination taking place and all the radicals formed are being removed on the oxide surface.

PROPYLENE

Before the technique of measuring the blueing rate of molybdenum oxide was developed as a measure of the rate of addition of hydrogen atoms to the surface, the problem of measuring the ratio of the number of atoms reaching the layer to the total number produced was tackled by measuring the number of double bonds removed in the gas phase by the technique already described involving the hydrogenation of a sample in the micro-hydrogenation gauge. This method could not be considered satisfactory since it involved removal of a finite amount of olefine and consequent uncertainty as to the correct concentration of olefine to be applied to the calculation. These results can be regarded only as giving the approximate order of magnitude of the collision efficiency.

The reaction vessel which was used was constructed from a filter funnel with a top plate of silica waxed in position and a glass plate fitted inside, 2 mm. from the top. The experiment was in the nature of an exploratory one to ascertain whether a short-path technique could be adopted or not.

The rate of production of atomic hydrogen was obtained by conducting a blank run with hydrogen. All the hydrogen atoms produced were removed on the molybdenum oxide surface and hence the drop in pressure of the hydrogen gave a measure of the rate of dissociation of the molecule hydrogen. For this determination, the reaction was opened to the manometer system. The ratio of the reaction vessel to the total volume of the reaction vessel and manometer was 1/1·9. The result obtained is shown in figure 4. From the slope, the average rate of removal of hydrogen is 0·2 mm./min. Now, if the available volume had been the reaction vessel alone, the rate of removal would have been 0·2 × 1·9 mm./min. = 0·38 mm./min.

The results obtained by carrying out runs with propylene are given in table 5. This table requires a few words of explanation. Column 2 gives the average pressure of propylene during the run. Column 3 gives the average pressure of hydrogen during the run. Column 5 gives the amount of propylene removed from the system as determined by unsaturate analysis before and after the run in the micro-hydrogenation gauge. Column 7 gives the amount of molecular hydrogen required to remove the propylene.

TABLE 5. RUNS WITH PROPYLENE—SPACING 2 MM.

run no.	$p_{C_3H_6}$ (mm.)	p_{H_2} (mm.)	time of irradiation (min.)	C_3H_6 used (mm.)	H_2 dissociated (mm.)	H_2 used by C_3H_6 (mm.)	H_2 on oxide (mm.)	R
145/1	0·8	5·4	1	0·5	0·38	0·25	0·13	0·34
145/2	0·7	5·2	1	0·47	0·38	0·24	0·14	0·37
146/3	0·75	5·6	1	0·34	0·38	0·17	0·11	0·29

Since the reaction is

$$H + C_3H_6 = C_3H_7,$$

followed by removal of the radicals, the pressure of molecular hydrogen required to remove the propylene is half the pressure of propylene removed.

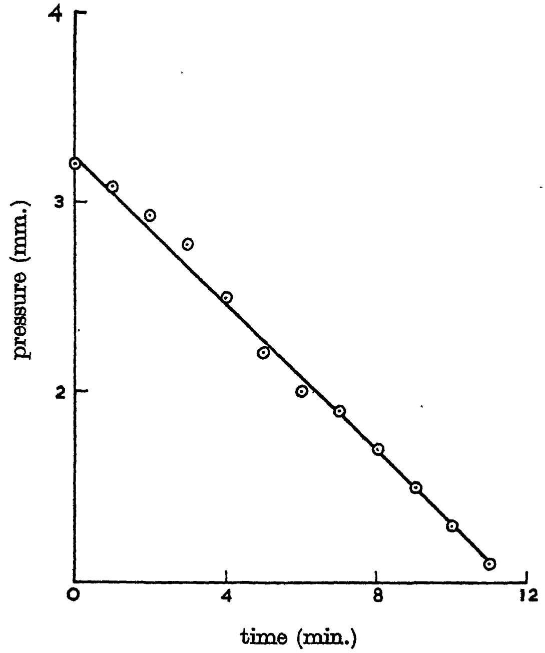


FIGURE 4. Removal of hydrogen atoms on molybdenum trioxide.

By subtracting column 7 from column 6, the amount of molecular hydrogen removed on the oxide layer to the total amount dissociated. This is the required ratio for determining the value of b from the curve given in part II, figure 4, for a spacing of 2 mm. The results obtained for this experiment are given in table 6. In this table, column 3 gives the value of b. Since by definition, $b = (g/D)^{\frac{1}{2}}$, g can be determined, knowing the value of D, tabulated in column 4. Further, since $g = k \times n$, as already defined, k can be evaluated and is recorded in column 6. k is given in the units ml.mol.$^{-1}$sec.$^{-1}$. Dividing k by the collision number for H-C_3H_6, the collision efficiency can be obtained.

TABLE 6. RESULTS OF RUNS WITH PROPYLENE

run no.	R	b	D	g	k	coll. eff.
145/1	0·34	12·6	$1{\cdot}09 \times 10^2$	$1{\cdot}7 \times 10^4$	6×10^{-13}	4×10^{-4}
145/2	0·37	11·5	$1{\cdot}16 \times 10^2$	$1{\cdot}5 \times 10^4$	6×10^{-13}	4×10^{-4}
145/3	0·29	14·1	$1{\cdot}1 \times 10^2$	$2{\cdot}0 \times 10^4$	$7{\cdot}5 \times 10^{-13}$	5×10^{-4}

In view of the limitations already discussed, these results can be expected only to give an approximate value for the collision efficiency of the reaction.

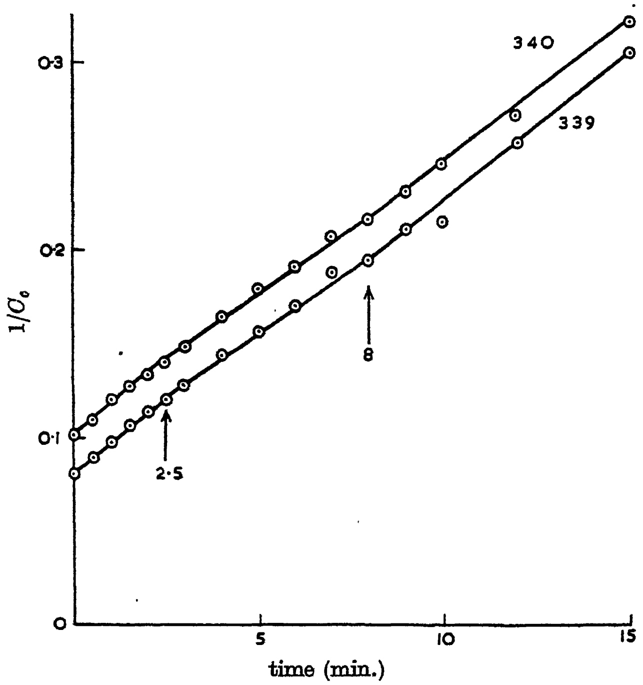


FIGURE 5. Runs with propylene—spacing 2·5 mm. (Tungsten oxide). 339 and 349: 6·0 mm. H_2, 0·5 mm. propylene.

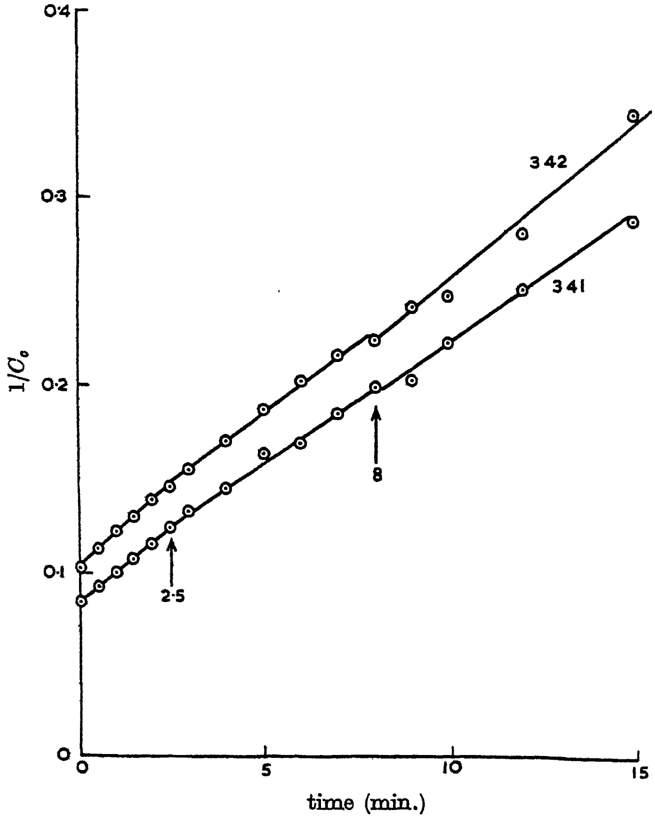


FIGURE 6. Runs with propylene—spacing 1·25 mm. (Tungsten oxide). 341 and 342: 6·0 mm. H_2, 0·5 mm. propylene.

The blueing technique has been applied to the determination of the collision efficiency of the reaction.

$$H + C_3H_6 = C_3H_7.$$

The corrected experimental curve for a separation of 2·5 mm. is shown in figure 5. Run 340 has been shifted upwards by 0·02 unit for clarity. The ratio of the slopes is 0·71 in both runs.

For a spacing of 1·25 mm. the corrected curves shown in figure 6 were obtained. Run 342 is shifted 0·02 unit upwards. The ratio for run 341 is 0·88 and for run 342 is 0·90. This gives an average of 0·89. Setting the values obtained for these two separations on the calculator, the value obtained for b is 6·0 and for k_R, 0·34.

The value of D is $1{\cdot}1 \times 10^2$ cm.²/sec., therefore,

$$g = b^2 . D = 3{\cdot}96 \times 10^3.$$

$$k = 3{\cdot}96 \times 10^3 / 0{\cdot}5 \times 3{\cdot}56 \times 10^{16} = 2{\cdot}2 \times 10^{-13}\,\text{ml.mol.}^{-1}\text{sec.}^{-1}.$$

Collision efficiency is then given by

$$2{\cdot}2 \times 10^{-13} / 1{\cdot}6 \times 10^{-9} = 1{\cdot}4 \times 10^{-4}.$$

Some criticism has been levelled by Bamford (1947) at the technique developed here. The suggestion put forward is that the observed decrease in blueing rate in presence of olefine is due to a reaction,

$$Hg + C_3H_7 = C_3H_7 . Hg,$$

which lowers the mercury concentration in the gas-phase and so decreases the rate of production of hydrogen atoms. However, the fact that the collision efficiency of this reaction has been determined by two distinct methods as described above with agreement as close as can reasonably be expected, indicates that any occurrence of the above reaction removing mercury atoms does not significantly affect the results.

iso-BUTENE

The reaction of hydrogen atoms with *iso*-butene has been investigated by the blueing technique in the same way. The runs obtained at a spacing of 2·5 mm. are given in figure 7. The value of the ratio of the two slopes for curve 311 is 1·52 and for run 316 is 1·44. This gives an average slope of 1·48.

For a spacing of 1·25 mm. the data obtained from runs 314 and 315 enable the corrected curves shown in figure 8 to be obtained. Run 315 has been shifted 0·01 unit downwards for clarity. For run 314, the ratio of the slopes is 1·23 and for run 315 the ratio is 1·31. Setting these values on the calculator, the value of b is 12 and k_R is 1·6. This gives a value for g of $1{\cdot}25 \times 10^4$.

From this $$k = 0{\cdot}7 \times 10^{-12}\,\text{c.c.mol.}^{-1}\text{sec.}^{-1},$$

and the collision efficiency is

$$0{\cdot}7 \times 10^{-12} / 1{\cdot}5 \times 10^{-9} = 4{\cdot}7 \times 10^{-4}.$$

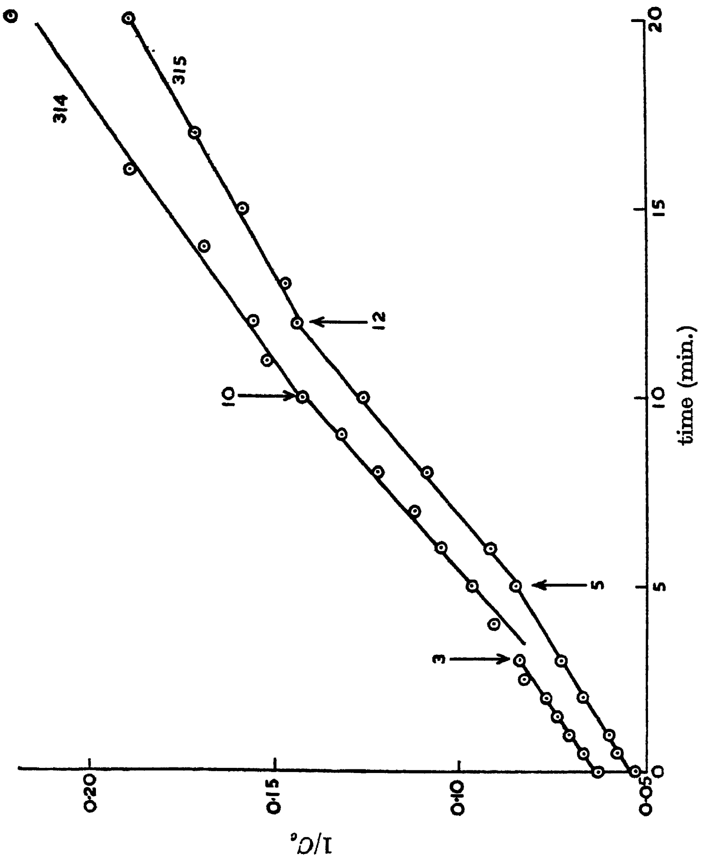


FIGURE 8. Runs with *iso*-butene—spacing 1·25 mm. (Tungsten oxide). 314: 7·8 mm. H_2, 0·5 mm. *iso*-butene. 315: 8·0 mm. H_2, 0·5 mm. *iso*-butene.

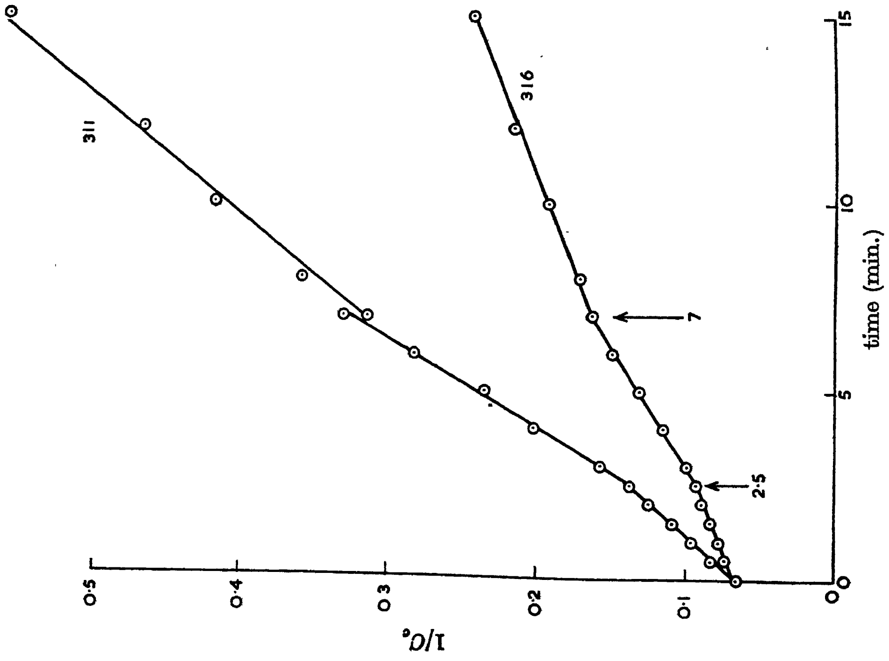


FIGURE 7. Runs with *iso*-butene—spacing 2·5 mm. (Tungsten oxide). 311: 7·6 mm. H_2, 0·5 mm. *iso*-butene. 316: 7·8 mm. H_2, 0·5 mm. *iso*-butene.

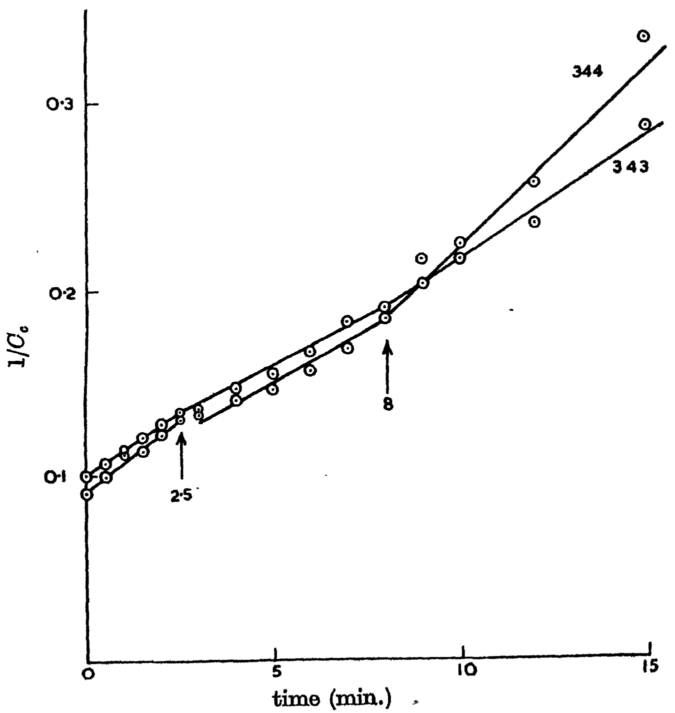


FIGURE 9. Runs with *n*-pentene-2 (*cis*)—spacing 2·5 mm. (Tungsten oxide). 343: 5·8 mm. H_2, 0·5 mm. *n*-pentene-2. 344: 6·0 mm. H_2, 0·5 mm. *n*-pentene-2.

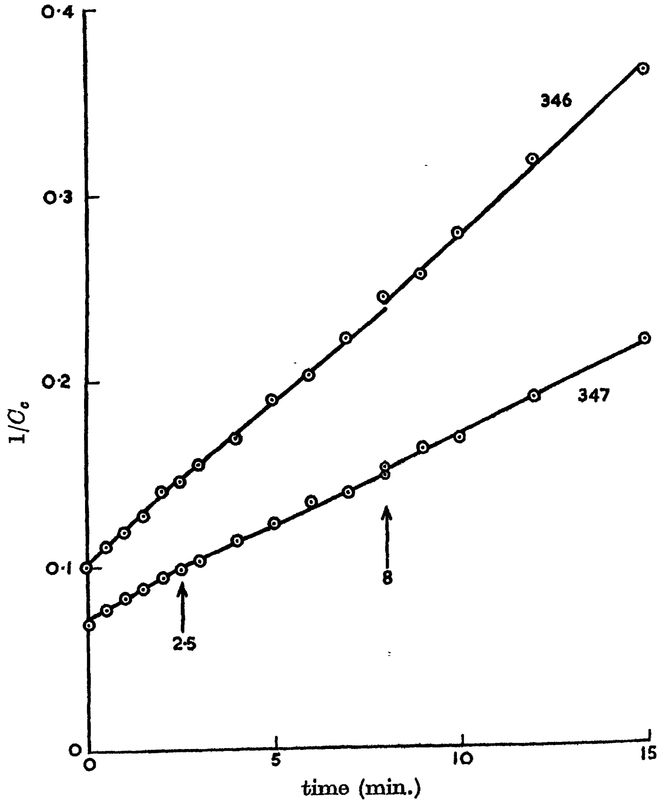


FIGURE 10. Runs with *n*-pentene-2 (*cis*)—spacing 1·25 mm. (Tungsten oxide). 346: 6·0 mm. H_2, 0·5 mm. *n*-pentene-2. 347: 6·1 mm. H_2, 0·5 mm. *n*-pentene-2.

n-PENTENE-2 (*cis*)

The results for a spacing of 2·5 mm. are given in figure 9. For run 343 the value of the ratio of the slopes is 0·83 and for run 344, the ratio is 0·78. This gives an average slope of 0·805.

For a spacing of 1·25 mm. runs 346 and 347 were obtained as shown in figure 10. For run 346, the ratio of the slopes is 0·90 and for run 347, the ratio is 0·89. This gives an average of 0·895.

From the calculator, the value of b is 11 and the value of k_R is 0·72. The diffusion coefficient D is $0{\cdot}85 \times 10^2$. This gives a value for g of $1{\cdot}03 \times 10^4$ and hence

$$k = 5{\cdot}8 \times 10^{-13}\,\text{ml.mol.}^{-1}\text{sec.}^{-1}.$$

The collision efficiency is then

$$5{\cdot}8 \times 10^{-13}/2{\cdot}1 \times 10^{-9} = 2{\cdot}8 \times 10^{-4}.$$

2, 3, 3, TRI-METHYL-BUTENE-1

This olefine proved to be most interesting since the radical produced had precisely the same blueing effect as had a hydrogen atom on tungsten oxide. Run 350 is typical of the behaviour with tungsten oxide, at a spacing of 2·5 mm. This run is shown in figure 11. It is readily appreciated that there is no change in the slope of the lines in the two parts of the graph.

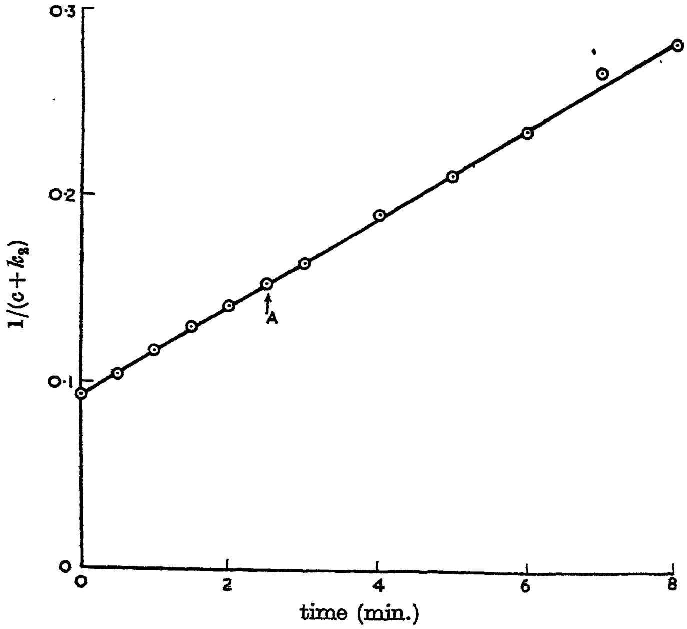


FIGURE 11. Run with 2, 3, 3, tri-methyl-butene-1 and tungsten oxide—spacing 2·5 mm. 6·0 mm. H_2 + 0·5 mm. olefine at *A*.

The removing agent was changed to molybdenum oxide to investigate whether or not the effect would remain the same. It was observed that with molybdenum oxide, the blueing effect did in fact alter.

The runs for a spacing of 2·5 mm. are given in figure 12. For run 351, the ratio of the slopes is 1·18 and for run 352, the ratio is 1·15. This gives an average of 1·165.

For a separation of 1·25 mm. the data obtained are plotted in figure 13. For run 353, the ratio of the slopes is 1·14 and for run 355, the ratio is 1·11. The average ratio is then 1·125.

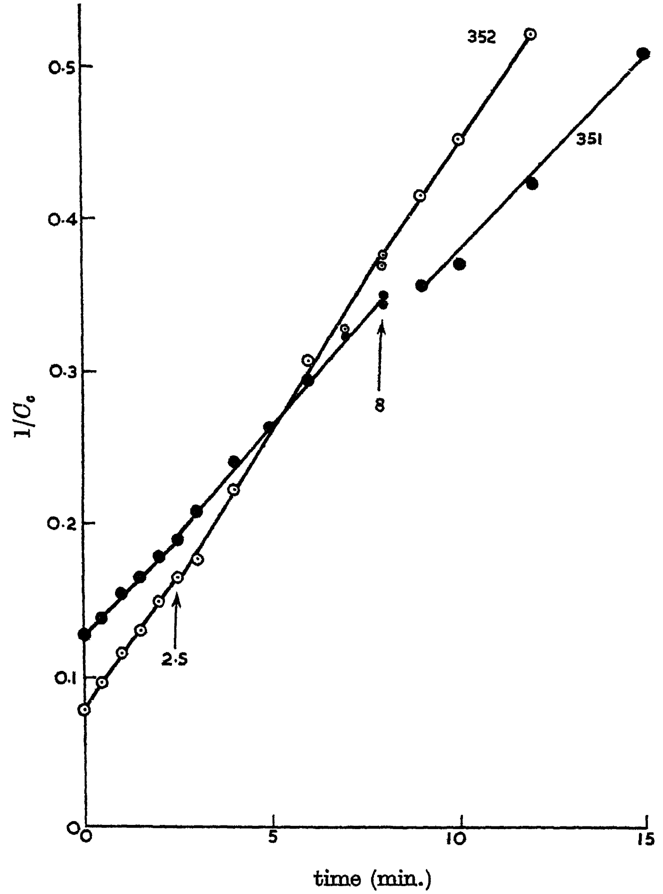


FIGURE 12. Runs with 2, 3, 3, tri-methyl-butene-1—spacing 2·5 mm. (Molybdenum oxide). 351 and 352: 6·0 mm. H_2, 0·5 mm. olefine.

By setting these values on the calculator, the value of b is difficult to ascertain. The closeness of the ratios, determined by the very similar effect of the radical and a hydrogen atom in blueing the surface precludes accurate determination of the value of b. Further, the low ratios are very sensitive to even slight adjustment of the slopes of the lines on the graph. It is judged, however, that the value of b lies between 10 and 20 and k_R has a value of about 1·2. This means that b^2 lies between 100 and 400. Taking the value of 200 for b^2,

$$g = 1{\cdot}76 \times 10^4.$$

This gives a value for k of

$$1{\cdot}76 \times 10^4/0{\cdot}5 \times 3{\cdot}56 \times 10^{16} = 9{\cdot}9 \times 10^{-13}\,\text{ml.mol.}^{-1}\text{sec.}^{-1}$$

which leads to a value for the collision efficiency of

$$5{\cdot}2 \times 10^{-4}.$$

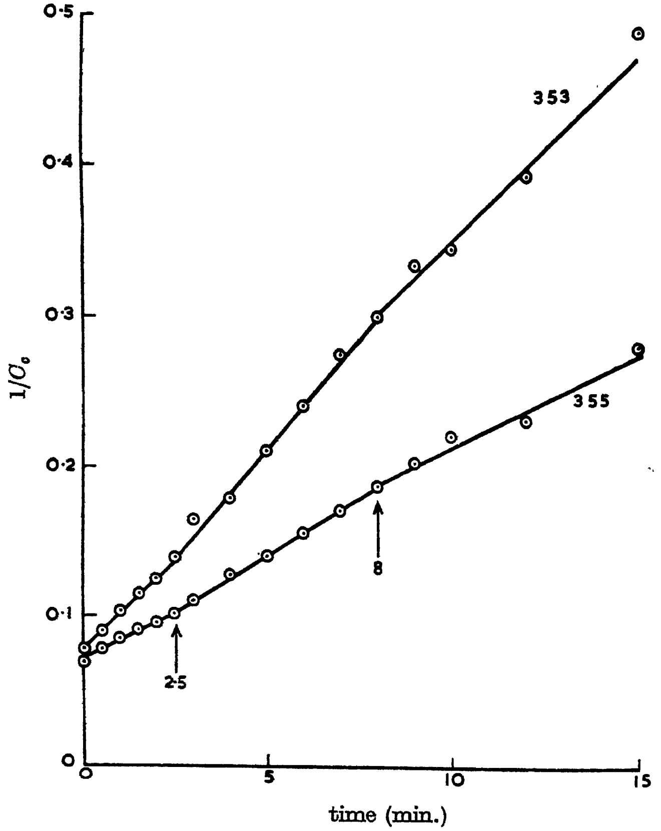


FIGURE 13. Runs with 2, 3, 3, tri-methyl-butene-1—spacing 1·25 mm. (Molybdenum oxide). 353 and 355: 6·0 mm. H_2, 0·5 mm. olefine.

SUMMARY OF RESULTS

In table 7 are collected the data obtained for the various olefines.

TABLE 7. COLLISION EFFICIENCY OF A HYDROGEN ATOM WITH OLEFINES

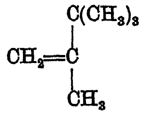
olefine	k	coll. eff.	k_R
ethylene	$9{\cdot}1 \times 10^{-13}$	$8{\cdot}5 \times 10^{-4}$	0·38
propylene	$2{\cdot}2 \times 10^{-13}$	$1{\cdot}4 \times 10^{-4}$	0·34
i-butene	$7{\cdot}0 \times 10^{-13}$	$4{\cdot}7 \times 10^{-4}$	1·6
n-pentene-2 (*cis*)	$5{\cdot}8 \times 10^{-13}$	$2{\cdot}8 \times 10^{-4}$	0·72
2, 3, 3, tri-methyl-butene-1	$9{\cdot}9 \times 10^{-13}$	$5{\cdot}2 \times 10^{-4}$	1·0

In this table, the velocity constant k and the collision efficiency of the reactions are listed. In addition, the blueing constant for the radial k_R, relative to a hydrogen atom is given on the surface of tungsten oxide. The velocity constant is expressed in ml.mol.$^{-1}$sec.$^{-1}$. All these observations were made at room temperature, between 14 and 18° C.

OBSERVATIONS

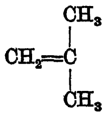
In the determination of the collision efficiency, the molecular diameter is involved. This does not seem to be very satisfactory since perhaps in the case of complex molecules, a more intimate part of the molecule is involved and the collision diameter should perhaps be a function of the close surroundings of the double bond. The errors involved will not be significant in dealing with the smaller molecules, but could become appreciable when larger molecules are concerned.

It is difficult to attempt generalizations on the data given above, but it may be tentatively suggested that the collision efficiency increases with the number of alkyl groups replacing the hydrogen atoms of ethylene. Ethylene does not fit into this scheme, but since it is not subject to the presence of bulky alkyl groups, this is perhaps not suprising. Propylene, with substitution of only one hydrogen atom is the least reactive. *n*-pentene-2, with replacement of two of the symmetrically disposed hydrogen atoms is next in order of reactivity, while *iso*-butene and 2, 3, 3, trimethyl-butene-1 with replacement of two unsymmetrical hydrogen atoms are next in order. Since there is little difference between the activities of the two latter olefines,



$CH_2{=}C(CH_3)C(CH_3)_3$

and



$CH_2{=}C(CH_3)_2$

it might be that the bulk size of the substituent, in the former an *iso*-butyl group and in the latter, a methyl group, has little effect on the reactivity, the mere presence of a group being sufficient to determine the reactivity.

These observations are merely speculative, but when many more data are accumulated, a true correlation of the effect of substitution on the reactivity of the olefinic bond may be obtained.

We wish to thank Dr A. J. Wildschut (Amsterdam), Bataafsche Petroleum Maatschappij for samples of the more uncommon olefines used in these investigations, and the Trustees of the Carnegie Trust for the award of a Scholarship to one of us (J.C.R.).

REFERENCE

Bamford, C. H. 1947 *Faraday Soc. Discussion*, 1947, 2.

Molecular orbitals and the Hartree field

By W. Moffitt, *New College, Oxford*

(*Communicated by F. E. Simon, F.R.S.—Received* 6 *August* 1948—
Revised 11 *November* 1948)

The aim of this paper has been to introduce self-consistency into a general, but necessarily rather oversimplified, method of molecular orbitals. Thus the non-linearity of the equations (19) has enabled us to deal in a systematic way with the charge distribution and bond properties of different conjugated molecules in a variety of configurations. Further analysis is required before term values can be predicted. The electro-affinity scale which may be set up for both σ and π electron pair bonds by means of this method turns out to be identical with that of Mulliken. A general account of the configurational theory of molecular structure has been given in an introductory section.

Introduction

The conventional molecular orbital method, which expresses one-electron molecular orbitals as linear combinations of atomic orbitals, has been successfully applied quantitatively to but a restricted class of molecules. And it is perhaps surprising at first sight that this class should contain only molecules with a reasonably uniform charge distribution, which are in general tractable by means of the HLSP or pairing approximation. Formally, the molecular orbital method is apparently much the better suited to treatments of heteropolar molecules—and, indeed, its qualitative success in the hands, particularly, of Mulliken, would lead one to hopes of its more general validity. It is proposed to apply the results of a detailed examination of this so-called LCAO approximation to the problem of the bonding between systems of atoms with different electronegativities. The larger part of our discussion will be confined to that of the unsaturation or π electrons of linear or planar molecules; the individual molecular orbitals will either have nodal surfaces in the molecular planes, or be characterized as having unit components of angular momentum about the molecular axes. It is believed that rather similar considerations may also be applied to σ electrons.

Atomic and molecular configurations

Owing to the similarity between the configurational theory of atomic spectra and the molecular orbital theory, it is proposed to introduce the latter briefly in the light of this analogy.

The configurational theory of atomic spectra begins with a set of one-electron functions $\phi^{(\alpha)}$, where (α) represents the fourfold symbol $(N^{(\alpha)}, l^{(\alpha)}, m_l^{(\alpha)}, m_s^{(\alpha)})$. These $\phi^{(\alpha)}$ are orthonormal inasmuch as

$$\int \phi^{(\alpha)*}\phi^{(\beta)}\,dv = \delta^{(\alpha)(\beta)}.$$

The azimuthal quantum number $l^{(\alpha)}$ indicates that irreducible representation of the atom's spatial symmetry group (three-dimensional rotation group) to which $\phi^{(\alpha}$

belongs, and the magnetic quantum number $m_l^{(\alpha)}$ designates some row of this $(2l^{(\alpha)}+1)$-fold degenerate representation. The principal quantum number $N^{(\alpha)}$ serves to distinguish the various $\phi^{(\alpha)}$ with the same transformation properties but with different 'energy parameters' $\epsilon^{(\alpha)}$, and $m_s^{(\alpha)}$ represents the spin co-ordinate. It is assumed that this set of functions contains no accidental degeneracies: $\epsilon^{(\beta)} = \epsilon^{(\alpha)}$ only holds when $N^{(\beta)} = N^{(\alpha)}$ and $l^{(\beta)} = l^{(\alpha)}$. For an n-electron system, the set of all product functions

$$\phi^{(1)}(l)\ldots\phi^{(\alpha)}(m)\ldots\phi^{(n)}(n),$$

which may be formed by choosing the $(\alpha) = (N^{(\alpha)}, l^{(\alpha)}, m_l^{(\alpha)}, m_s^{(\alpha)})$ so as to be consistent with Pauli's exclusion principle and a given total zeroth-order energy $\epsilon = \sum_{\alpha=1}^{n} \epsilon^{(\alpha)}$, and by permuting the electrons m amongst the $\phi^{(\alpha)}$, may be called a configuration. Since ϵ depends only on the $N^{(\alpha)}$ and $l^{(\alpha)}$, a given configuration may be uniquely specified more simply by the notation

$$\ldots[\phi^{\beta}]^{b}\ldots[\phi^{\alpha}]^{a}\ldots \quad (\alpha = 1, \ldots, \mu);$$

α now only represents the twofold symbol (N^{α}, l^{α}) and $a = n^{\alpha}$ is the number of electrons in this configuration which are associated with the functions ϕ^{α}, so that $n = \sum_{\alpha=1}^{\mu} n^{\alpha}$. The spectral states and their term symbols which arise from these configurations may easily be found (e.g. Slater 1929), and give a satisfactory account of the essential features of atomic spectra.

The best one-electron atomic orbitals ϕ^{α} may be obtained as central, but non-Coulomb, field functions by Hartree's method (1928)—in which each electron is supposed to move in the field of the nucleus and the time-average field of all the other electrons. The energy parameters ϵ^{α} which occur in the differential equations for the ϕ^{α}, obtained by separating the many-electron Schrödinger equation in this way, are found to give surprisingly good values for the ionization potentials corresponding to the removal of electrons in the ϕ^{α}. Further, the electronic charge distribution and magnetic properties of atoms and ions may be calculated directly from these so-called self-consistent field functions.

In order to obtain a more detailed account of atomic spectra including Heisenberg resonance effects, it is necessary to utilize the full-determinantal (anti-symmetrized) many-electron functions which may be constructed for any state from the product functions characterizing the parent configuration. This may be readily accomplished using either Slater's method (1929) or Dirac's vector model of the atom (van Vleck 1934*b*).

Now, just as in the case of atoms, electronic configurations may be assigned also to molecules:

$$\ldots[\psi^{\rho}]^{r}\ldots[\psi^{\tau}]^{t}\ldots \quad (\tau = 1, \ldots, \nu).$$

In order that these be significant it is necessary that the one-electron molecular orbitals ψ^{τ} be orthonormal

$$\int \psi^{\rho *}\psi^{\tau}\,dv = \delta^{\rho\tau},$$

and further, that they belong to irreducible representations of the molecule's spatial symmetry group (point group). τ, just as α for atoms, is a twofold symbol. It indicates, first, the transformation properties or symmetry characteristics of ψ^{τ}—as did l^{α} for the ϕ^{α}—and secondly, it contains an ordering symbol, analogous to N^{α} in the atomic case, which distinguishes molecular orbitals belonging to the same irreducible representation but with different energy parameters or 'vertical ionization potentials'. $t = n^{\tau}$ is the number of electrons of the n-electron system which are associated with ψ^{τ}, so that $n = \sum_{\tau=1}^{\nu} n^{\tau}$. If the representation to which ψ^{τ} belongs is g^{τ}-fold degenerate, it is clear that $n^{\tau} \leqslant 2g^{\tau}$, by Pauli's principle; similarly, in the case of atoms, of course, $a = n^{\alpha} \leqslant 2(2l^{\alpha}+1)$, and for the same reason.

In constructing atomic configurations it is usual to begin using simplified one-electron atomic orbitals (e.g. Slater 1930)—and these are generally a sufficient guide to an explanation of the qualitative features of atomic spectra. In the case of molecules, one may begin in one of two different ways. For certain diatomic molecules and very simple polyatomic molecules, such as methane, it may be convenient to use the united-atom approach, but a more generally useful method is that of LCAO molecular orbitals which was developed by Hund (1933) and Mulliken (1934*a*). LCAO molecular orbitals are functions ψ^{τ} which are constructed as linear combinations of atomic orbitals ϕ_i^{τ}. (The subscript i refers to the atom with which ϕ_i^{τ} is associated; the superscript τ characterizes that orbital of atom i which is to be a constituent of ψ^{τ}; τ, for the ϕ_i^{τ}, contains not only the symbol (N_i^{τ}, l_i^{τ}), but also includes a descriptive 'local' symmetry symbol and an appropriate screening constant.)

The LCAO molecular orbitals and the electronic configurations which may be constructed from them have been conspicuously successful in explaining qualitatively the ultra-violet spectra of molecules, particularly of those with high symmetry (e.g. Mulliken 1934*a*). For the molecular states arising from any given configuration are easily written down together with their molecular term symbols.

For the best ψ^{τ}'s one would presumably again use Hartree field function; the loss of spherical symmetry, however, introduces prohibitive numerical difficulties which make these but distant ideals. Accordingly, if quantitative results are to be obtained at all, rather drastic approximations must be introduced, and the most common of these is to use those LCAO forms which minimize the zeroth-order energy of the configuration. Unfortunately, this method can, at present, only be carried through plausibly for certain conjugated systems with reasonably uniform charge distributions (e.g. Coulson & Longuet-Higgins (1947), where back references will be found); in other cases predictions tend to depend rather critically on an appropriate choice of rather too many 'semi-empirical parameters'. It is the aim of this paper to examine in greater detail the foundations of the LCAO approximation, and to suggest a more general method for its application to a wider range of molecular types. We shall, therefore, develop an LCAO approximation, primarily for π electrons, which minimizes the zeroth-order energy; the overlap integrals are 'transformed away' in order to simplify the discussion.

From a knowledge of the best LCAO forms of the ψ^{τ} it should be possible to calculate the electronic charge distribution in molecules, to predict their magnetic and bond properties and to estimate the 'vertical ionization potentials' of their various electrons. In order to explain the more quantitative aspects of ultra-violet spectra —just as for atoms—antisymmetrization is necessary (Goeppert-Mayer & Sklar 1938) and may be attained most simply by an adaptation of Dirac's vector model to molecules. The intensities of electronic spectra have, however, been calculated relatively successfully (Mulliken 1939) using simple LCAO ψ^τ's without this complication.

THE LCAO MOLECULAR ORBITALS

Consider a configuration of $n = \sum_\tau n^\tau$ mobile π electrons in ν different molecular orbitals ψ^τ, each containing $t = n^\tau$ electrons:

$$\ldots[\psi^\rho]^r\ldots[\psi^\tau]^t\ldots \quad (\tau = 1,\ldots,\rho,\ldots,\nu).$$

The LCAO forms for these molecular orbitals, which are responsible for the π-bonding between N atomic nuclei, will be

$$\psi^\tau = \sum_i \gamma_i^\tau \phi_i^\tau \quad (i = 1,\ldots,N); \tag{1}$$

here ϕ_i^τ refers to the best atomic orbital of nucleus i for use in constructing ψ^τ and γ_i^τ is its coefficient in ψ^τ. In general, as is implied in our notation, a slightly different atomic function from any given atom is required for each molecular orbital, corresponding to a different screening constant Z_i^τ in each case. It will always be possible, for our purposes, to arrange for the ϕ_i^τ and γ_i^τ to be real, and we shall assume that this is indeed the case.

If $\mathscr{H}^\tau$ is the self-consistent Hamiltonian operator which determines the motion of an electron in the molecular orbital ψ^τ, it may be written in the form

$$\mathscr{H}^\tau = -\frac{h^2}{8\pi^2 m}\nabla^2 + \sum_i v_i^\tau(r_i), \tag{2}$$

where $v_i^\tau(r_i)$ represents the contribution from the ith attracting centre to the total potential energy $V^\tau = \sum_i v_i^\tau$; and V^τ consists of the potential energy due to the nuclear skeleton and to the self-consistent field of all the other electrons. Ideally, therefore, the atomic orbitals would be solutions of the equations

$$\left\{-\frac{h^2}{8\pi^2 m}\nabla^2 + v_i^\tau(r_i)\right\}\phi_i^\tau(r_i) = \eta_i^\tau \phi_i^\tau(r_i), \tag{3}$$

and we might call $-\eta_i^\tau$ the 'term value' of an electron in ϕ_i^τ. However, the decomposition $\mathscr{V}^\tau = \sum_i v_i^\tau$ is not unique, since we have no unambiguous way of assigning the self-consistent field of the other electrons to the various nuclei. In regions where v_i^τ is large, for example, v_j^τ will be small but will not, in general, vanish (cf. (*d*), below). Thus the v_i^τ are not defined by means of fields due to electronic distributions within closed surfaces about the atoms i, but will depend in a much more complicated fashion on the solutions $\phi_i^\tau(r_i)$ of the equations (3) with which they are associated.

Fortunately, we do not need to specify the ϕ_i^τ explicitly, and it will be sufficient that the v^τ satisfy the following requirements:

(*a*)
$$\sum_i v_i^\tau = \mathscr{V}^\tau.$$

(*b*) It is necessary that the ϕ_i^τ should not lose their property of being essentially atomic orbitals. Therefore, as the atoms go separately to infinity while retaining the charge associated with them in the molecule (*vide infra*), so v_i^τ must become $\mathbf{v}_i$—which is the field of the nucleus i, the non-π-bonding electrons of i and the other π valence electrons associated with i. ϕ_i^τ then becomes $\boldsymbol{\phi}_i^\tau$, satisfying

$$\left\{-\frac{h^2}{8\pi^2 m}\nabla^2 + \mathbf{v}_i^\tau(r_i)\right\}\boldsymbol{\phi}_i^\tau(r_i) = \boldsymbol{\eta}_i^\tau\,\boldsymbol{\phi}_i^\tau(r_i), \tag{4}$$

and is a proper atomic 'valence state' function whose screening constant $\mathbf{Z}_i^\tau$ is determined by $\mathbf{v}_i^\tau$ and whose term value is $-\boldsymbol{\eta}_i^\tau$.

(*c*) The v_i^τ transform, under the operations of the molecule's symmetry group, like the atomic nuclei to which they refer, i.e. they have the symmetry of the corresponding atomic sites in the molecule.

(*d*) The v_i^τ remain as close to the $\mathbf{v}_i^\tau$ as is consistent with the condition that the orthonormality of the ϕ_i^τ may be expressed, to some arbitrary degree of approximation, by

$$\int \psi^\tau \psi^\rho \, dv = \sum_i \gamma_i^\tau \gamma_i^\rho = \delta^{\tau\rho} \quad (\text{all } \rho, \tau). \tag{5}$$

(The condition (*d*) may be replaced at a later stage of the argument by a neglect of the overlap integrals between atomic functions. For a semi-empirical method such as the LCAO method, however, our procedure is logically if not fundamentally more satisfactory.) We notice that, in general, the ϕ_i^τ will no longer be central field functions, but will be closely related to the $\boldsymbol{\phi}_i^\tau$, which may be taken as central, but only roughly as Coulomb, field functions.

BOND ORDERS AND CHARGE DISTRIBUTION

Let us assume that the coefficients γ_i^τ have been determined by a variational treatment so that we know the forms of each of the molecular orbitals

$$\psi^\tau = \sum_i \gamma_i^\tau \phi_i^\tau.$$

By normalizing (5) we shall have

$$\sum_i (\gamma_i^\tau)^2 = 1,$$

so that it is natural to associate some $(\gamma_i^\tau)^2$ of the charge of an electron in ψ^τ with atom i. The net formal charge associated with this atom is therefore (in units of the electronic charge)

$$Q_i = z_i - q_i = z_i - \sum_\tau n^\tau (\gamma_i^\tau)^2, \tag{6}$$

where z_i is the charge on the atom i when stripped of its π electrons. Similarly (Coulson 1939) the mobile bond order between neighbouring atoms i and j may be considered as made up of contributions $n^\tau \gamma_i^\tau \gamma_j^\tau$ from each molecular orbital shell $[\psi^\tau]^{n^\tau}$:

$$p_{ij} = \sum n^\tau \gamma_i^\tau \gamma_j^\tau. \tag{7}$$

It is primarily with these quantities q_i, p_{ij} that Coulson & Longuet-Higgins (1947) have associated the properties of conjugated systems, though their method of determining the γ_i^τ differs from that which we shall develop.

We may now examine η_i^τ and v_i^τ of (4) more closely. v_i^τ has been defined as the field determining the motion of an electron in ψ^τ around atom i, when the atoms are separated while retaining the charges associated with them in the molecule. That is, v_i^τ is the field of the nucleus i, the non-π-bonding electrons of i and the remaining

$$\omega_i^\tau = \sum_\rho (n^\rho - \delta^{\rho\tau})(\gamma_i^\rho)^2 \tag{8}$$

electrons associated with the π valence shell of this atom; $-\eta_i^\tau$ is the term value of an electron in such a field. When there is only $n = 1$ π-bonding electron in the molecule and $z_i = 1$, it is clear that $\omega_i^\tau = 0$ and $-\eta_i^\tau$ may be taken as the ionization potential I_i of a π electron from the appropriate valence state of the atom i. Similarly, if $z_i = 1$ and $\omega_i^\tau = 1$, $-\eta_i^\tau$ is the corresponding electron affinity E_i. In general, therefore,

$$\eta_i^\tau = \eta_i(\omega_i^\tau) \tag{9}$$

is a function of ω_i^τ which, when ω_i^τ assumes integral values, may be related to spectroscopic data on atom i.

We note that I_i and E_i refer not to processes in which actual spectroscopic states of atoms are involved, but atomic 'valence states' (van Vleck 1934*a*). A valence state of an atom is chosen (and mathematically defined) so as to have as nearly as possible the same intra-atomic electronic spin coupling as when the atom is part of a molecule. Its term values may be fairly accurately related to those of the spectroscopic states by means of certain interaction integrals introduced by Slater (1929). Mulliken (1934*b*) has calculated such ionization potentials and electron affinities for many atoms in various valence states.

We shall assume that for non-integral values of ω_i^τ, the appropriate η_i^τ's may be found by some interpolation between their values which may be derived from spectroscopic data. In many cases where $z_i = 1$, it will be shown, for example, that it is reasonable to take the linear formula

$$\eta_i^\tau = \eta_i(\omega_i^\tau) = -I_i + (I_i - E_i)\,\omega_i^\tau. \tag{10}$$

THE VARIATION METHOD

The LCAO method proceeds by minimizing the zeroth-order energy of the molecular orbital configuration with respect to the coefficient-parameters γ_i^τ. The total zeroth-order energy is

$$\eta = \sum_\tau n^\tau \eta^\tau, \quad \eta^\tau = \int \psi^\tau \mathscr{H}^\tau \psi^\tau \, dv,$$

or, in the approximation (1) that the molecular orbitals are linear combinations of the atomic orbitals ϕ_i^τ,

$$\eta = \sum_\tau n^\tau \left[\sum_i \left\{ (\gamma_i^\tau)^2 \epsilon_i^\tau + \sum_j{}' (\gamma_i^\tau \gamma_j^\tau)\, \beta_{ij}^\tau \right\} \right] \quad (j \neq i); \tag{11}$$

using equations (2) and (3),

$$\epsilon_i^\tau = \int \phi_i^\tau \mathscr{H}^\tau \phi_i^\tau \, dv = \eta_i^\tau + \sum_j{}' \int \phi_i^\tau v_j^\tau \phi_i^\tau \, dv$$

and

$$\beta_{ij}^\tau = \int \phi_i^\tau \mathscr{H}^\tau \phi_j^\tau \, dv = \sum_k{}' \int \phi_i^\tau v_k^\tau \phi_j^\tau \, dv \quad (k \neq j)$$

are the so-called Coulomb and resonance integrals respectively. In performing this process of minimization, we must ensure that the molecular orbitals ψ^τ remain orthonormal, namely, that conditions (5) are satisfied

$$\sum_i \gamma_i^\tau \gamma_i^\rho = \delta^{\tau\rho}.$$

In order to carry through the analysis, certain simplifying assumptions will now have to be made. This is necessary because the functional forms of ϵ_i^τ and β_{ij}^τ, which will depend in no simple manner on the γ_i^τ and the equilibrium internuclear distances r_{ij}, are unknown. These assumptions have been chosen so as to exhibit the maximum number of physically significant factors consistent with relative analytical simplicity. In particular the compromise between ionic and covalent bonding, and the necessity of using self-consistent Coulomb integrals are included, while the usual properties of calculations on conjugation problems are retained. The eventual justification of any such assumptions will be in the measure of agreement attained between predicted and observed properties of molecules. No attempt will therefore be made to evaluate the integrals ϵ_i^τ, β_{ij}^τ; a zeroth-order treatment is very probably too approximate to warrant the attack of such quadratures in the first place. (It is because of this that we have introduced the polarized atomic functions ϕ_i^τ, e.g. (5), which simplify the formal development of our approximation, but which would complicate considerably the direct evaluation of these integrals.)

We shall assume, first, that the resonance integrals depend only on the nature of atoms i and j, so that $\beta_{ij}^\tau = \beta_{ij}$ for all τ. (This assumption would, however, have to be modified if both $p\pi$ and $d\pi$ electrons of the same atom were considered as participating in the conjugation—as is possibly the case for the sulphur atom in thiophene, for example.) Since the ϵ_i^τ will occur only as differences (*vide infra* for CO, $A\,^1\Pi$), our suppositions concerning these will be relative only: it would be rash to use these ϵ_i^τ's for the prediction of 'vertical ionization potentials' or orbital term values without further analysis. Our second assumption is now that

$$\epsilon_i^\tau = \boldsymbol{\eta}_i(\omega_i^\tau), \quad \omega_i^\tau = \sum_\rho (n^\rho - \delta^{\rho\tau})(\gamma_i^\rho)^2, \tag{12$'$}$$

where the function $\boldsymbol{\eta}_i$ is the same as that defined in equation (9) above and, more specifically, takes a linear form

$$\boldsymbol{\eta}_i(x) = \eta_i^0 + \dot{\eta}_i(x). \tag{13}$$

To a first approximation, therefore, it is assumed that, with an appropriate choice for the β_{ij}, we may take $\epsilon_i^\tau = \boldsymbol{\eta}_i^\tau$. This is certainly the simplest of feasible forms for the ϵ_i^τ when dealing with atoms of different electro-affinities; it is the more plausible when it is shown (*vide infra*) that this choice leads to an absolute electro-affinity scale which is identical with that set up so successfully on other theoretical grounds by Mulliken in 1934. If, however, we were to introduce these values in (11), we should count the interactions between the various electrons twice. Accordingly, with the linear form (13) for the functions $\eta_i(x)$, it is clear that when considering the total energy of a configuration we must use not the values (12)′ but rather

$$\epsilon_i^\tau = \eta_i(\tfrac{1}{2}\omega_i^\tau). \tag{12}$$

If we could treat all the γ_i^τ as independent variables we should have, using (8), (9), (12) and (13),

$$\frac{1}{2}\frac{\partial\eta}{\partial\gamma_i^\tau} = n^\tau\left\{\gamma_i^\tau\left(\epsilon_i^\tau + \tfrac{1}{2}\gamma_i^\tau\frac{\partial\epsilon_i^\tau}{\partial\gamma_i^\tau}\right) + \sum_j{}' \gamma_j^\tau\beta_{ij}\right\} + \sum_\rho{}' n^\rho(\gamma_i^\rho)^2\frac{1}{2}\frac{\partial\epsilon_i^\rho}{\partial\gamma_i^\tau} \quad (i\neq j;\ \rho\neq\tau),$$

or

$$\frac{1}{2}\frac{\partial\eta}{\partial\gamma_i^\tau} = n^\tau\left(\gamma_i^\tau\xi_i^\tau + \sum_j{}' \gamma_j^\tau\beta_{ij}\right), \tag{14}$$

where

$$\xi_i^\tau = \eta_i^\tau = \eta_i(\omega_i^\tau), \tag{15}$$

and thus

$$\tfrac{1}{2}\delta\eta = \sum_\tau\sum_i n^\tau(\gamma_i^\tau\xi_i^\tau + \sum_j{}' \gamma_j^\tau\beta_{ij})\,\delta\gamma_i^\tau. \tag{16}$$

However, the $n\nu$ coefficients γ_i^τ must satisfy the orthonormality conditions (5), so that the $\frac{1}{2}\nu(\nu+1)$ conditions

$$\delta\left(\sum_i \gamma_i^\tau\gamma_i^\rho\right) = \sum_i \gamma_i^\tau\delta\gamma_i^\rho + \sum_i \gamma_i^\rho\delta\gamma_i^\tau = 0 \tag{17}$$

must hold at the outset. Accordingly, we combine (16) and (17) by Lagrange's method of undetermined multipliers and obtain for our condition that η be a minimum

$$\sum_\tau\sum_i\left\{n^\tau\left(\gamma_i^\tau\xi_i^\tau + \sum_j{}' \gamma_j^\tau\beta_{ij}\right) - \sum_\rho \lambda^{\tau\rho}\gamma_i^\rho\right\}\delta\gamma_i^\tau = 0. \tag{18}$$

Here $[\lambda^{\tau\rho}]$ is a symmetric matrix. Since the $\delta\gamma_i^\tau$ are arbitrary, we obtain the equations ($n\nu$ in number)

$$n^\tau\left(\gamma_i^\tau\xi_i^\tau + \sum_j{}' \gamma_j^\tau\beta_{ij}\right) - \sum_\rho \lambda^{\tau\rho}\gamma_i^\rho = 0. \tag{19}$$

From these we may eliminate the $\frac{1}{2}\nu(\nu+1)$ undetermined multipliers $\lambda^{\tau\rho} = \lambda^{\rho\tau}$ and, together with the $\frac{1}{2}\nu(\nu+1)$ conditions (5), we shall have sufficient equations to determine our $n\nu$ unknown coefficients γ_i^τ. We notice that since ξ_i^τ has been defined by means of (15), the equations (19) are not, in general, linear. They are inevitably more complicated than those obtained by the usual secular equation method (e.g. Coulson & Longuet-Higgins 1947), but may be solved numerically by the method of steepest descents. Since it is generally possible to construct an approximate set of γ_i^τ, on qualitative grounds, no difficulties with multiple roots should be encountered.

AN EXAMPLE: CO, $A\,^1\Pi$

In order to illustrate our use of equations (19), consider as an example the excited state $A\,^1\Pi$ of carbon monoxide. This will be discussed more fully in the following paper (Moffitt 1949), some of whose conclusions we shall anticipate here. CO, $A\,^1\Pi$

has five π electrons and may be treated as arising from the configuration $[\psi]^4[\psi']$, where ψ, ψ' are both doubly degenerate, the former being a bonding and the latter an antibonding function. It is assumed that the σ electronic charge distribution is predominantly homopolar and may be neglected in the variational problem. In LCAO approximation, therefore,

$$\psi = \gamma_C\phi_C + \gamma_O\phi_O, \quad \psi' = \gamma'_C\phi'_C + \gamma'_O\phi'_O.$$

By orthonormality we have

$$\gamma_C^2 + \gamma_O^2 = 1 = \gamma_C'^2 + \gamma_O'^2, \quad \gamma_C\gamma'_C + \gamma_O\gamma'_O = 0,$$

and the equations (19) may in this case be written as

$$4(\gamma_C\xi_C + \gamma_O\beta_{CO}) + \lambda\gamma_C + \lambda'\gamma'_C = 0, \quad (\gamma'_C\xi'_C + \gamma'_O\beta_{CO}) + \lambda''\gamma'_C + \lambda'\gamma_C = 0,$$
$$4(\gamma_O\xi_O + \gamma_C\beta_{CO}) + \lambda\gamma_O + \lambda'\gamma'_O = 0, \quad (\gamma'_O\xi'_O + \gamma'_C\beta_{CO}) + \lambda''\gamma'_O + \lambda'\gamma_O = 0.$$

On eliminating λ, λ' and λ'', it is found that

$$4\gamma_C\gamma_O(\xi_C - \xi_O) + \gamma'_C\gamma'_O(\xi'_C - \xi'_O) + 4(\gamma_O^2 - \gamma_C^2)\beta_{CO} + (\gamma_O'^2 - \gamma_C'^2)\beta_{CO} = 0,$$

from which we may further eliminate γ_O, γ'_C and γ'_O by means of the orthonormality conditions and obtain the equation

$$\gamma_C\sqrt{(1-\gamma_C^2)}\{4(\xi_C - \xi_O) - (\xi'_C - \xi'_O)\} + 3\beta_{CO}(1 - 2\gamma_C^2) = 0. \tag{A}$$

Since the five π electrons may be considered as responsible for the bonding between the two nuclei C^{++} and O^{+++}, it is readily shown that

$$\begin{aligned}
\xi_C &= -I_C + (I_C - E_C)(3\gamma_C^2 + \gamma_C'^2 - 1) = -I_C + (I_C - E_C)(2\gamma_C^2),\\
\xi'_O &= -I_O + (I_O - E_O)(3\gamma_O^2 + \gamma_O'^2 - 2) = -I_O + (I_O - E_O)(1 - 2\gamma_C^2),\\
\xi'_C &= -I_C + (I_C - E_C)(4\gamma_C^2 - 1),\\
\xi'_O &= -I_O + (I_O - E_O)(4\gamma_O^2 - 2) \qquad = -I_O + (I_O - E_O)(2 - 4\gamma_C^2).
\end{aligned}$$

Accordingly these values may be substituted in (A) which becomes an equation containing only one unknown, namely, γ_C. When the appropriate root $(0 < \gamma_C < 1)$ has been extracted, γ_O, γ'_C and γ'_O are readily evaluated, so that the charge distribution and bond order in the molecule may be calculated from equations (6) and (7).

SOME SYMMETRY CONSIDERATIONS

It is interesting to find under what conditions the system of equations (19) reduces to the simpler and (*vide infra*) apparently conventional set

$$\gamma_i^\tau\left(\xi_i^\tau - \frac{1}{n^\tau}\lambda^{\tau\tau}\right) + \sum_j{}' \gamma_j^\tau\beta_{ij} = 0 \tag{20}$$

(and incidentally to evaluate the $\lambda^{\tau\rho}$). In order that this should occur, it is both necessary and sufficient that the matrix $[\lambda^{\tau\rho}]$ be diagonal. For if $\sum_\rho{}' \lambda^{\tau\rho}\gamma_i^\rho = \Delta_i^\tau$, then $\sum_i (\Delta_i^\tau)^2 = \sum_\rho{}' (\lambda^{\tau\rho})^2$ by conditions (5), so that if all the Δ_i^τ are to vanish so must all the $\lambda^{\tau\rho}$ with $\rho \neq \tau$. Let us therefore determine the $\lambda^{\tau\rho}$. On multiplying (19) by λ_i^ρ and summing over all i, using (5), it is found that

$$\lambda^{\tau\rho} = n^\tau \sum_i \left\{ (\gamma_i^\rho \gamma_i^\tau)\, \xi_i^\tau + \sum_j{}' (\gamma_i^\rho \gamma_j^\tau)\, \beta_{ij} \right\}. \tag{21}$$

Now it is always possible to find a set of functions θ_i^τ with the symmetry properties of the ϕ_i^τ, and a set of symmetrical operators $\mathscr{L}^\tau$ (symmetrical in the sense that they transform under the identical representation of the molecule's point group), such that

$$\int \theta_i^\tau \mathscr{L}^\tau \theta_i^\tau \, dv = \xi_i^\tau \quad \text{and} \quad \int \theta_i^\tau \mathscr{L}^\tau \theta_j^\tau \, dv = \beta_{ij}.$$

Then
$$\zeta^{\rho(\tau)} = \sum_i \gamma_i^\rho \theta_i^\tau, \quad \zeta^\tau = \sum_i \gamma_i^\tau \theta_i^\tau$$

transform like ψ^ρ, ψ^τ respectively, and (21) becomes

$$\lambda^{\tau\rho} = n^\tau \int \zeta^{\rho(\tau)} \mathscr{L}^\tau \zeta^\tau \, dv.$$

Accordingly, if and, in general, only if ψ^ρ, ψ^τ belong to different irreducible representations of the appropriate symmetry group, will $\lambda^{\tau\rho} = \lambda^{\rho\tau}$ vanish. In cases of high symmetry, therefore, our system of equations may be considerably simplified. The extremely simple form (20) of our equations (19) is only assumed when all the ν molecular orbitals ψ^τ transform differently—as, for example, in benzene and the ground state of glyoxal—i.e. when the orthogonality requirement is satisfied automatically.

Our evaluation of the $\lambda^{\tau\rho}$ of (19) by means of (5) suggests an alternative method of solving our equations (19). In many cases the elimination of the $\lambda^{\tau\rho}$ will be tedious algebraically, so that it is better to substitute the values (21) for these and then proceed with the numerical solution of the modified and symmetrical set of equations

$$\left(\gamma_i^\tau \xi_i^\tau + \sum_j{}' \gamma_j^\tau \beta_{ij}\right) - \sum_\rho \sum_l \left\{ (\gamma_l^\rho \gamma_l^\tau)\, \xi_l^\tau + \sum_k{}' (\gamma_l^\rho \gamma_k^\tau)\, \beta_{lk} \right\} \gamma_i^\rho = 0 \quad (k \neq l). \tag{22}$$

There are occasions when the variational method which we have outlined breaks down. Suppose we have a configuration

$$\ldots [\psi^\rho]^x \ldots [\psi^\tau]^x \ldots \quad (\tau = 1, \ldots, \nu = n),$$

where the number of independent molecular orbitals is equal to the number of atoms participating in the conjugation and $x \leqslant 2g$, each ψ^τ being g-fold degenerate. Such configurations would describe molecules which are in certain excited sub-Rydberg states (Mulliken 1939)—when $x < 2g$—or whose π valence shell is completely filled ($x = 2g$). It may readily be shown that, in our approximation, the π bond orders all vanish and the resultant electron distribution is uniform, whatever the nature of the constituent atoms: For we may regard the γ_i^τ as elements of a real square matrix γ, in general non-singular. Then the orthonormality condition

$$\sum_i (\gamma_i^\tau \gamma_i^\rho) = \delta^{\tau\rho}$$

expresses the condition that $\gamma\gamma' = 1$, where γ' is the transpose of γ. Consequently γ is real orthogonal and we have

$$\sum_{\tau} (\gamma_i^{\tau} \gamma_j^{\tau}) = \delta_{ij}.$$

But for such configurations our equations in the form (22) reduce to the identity

$$\gamma_i^{\tau} \xi_i^{\tau} + \sum_j{}' \gamma_j^{\tau} \beta_{ij} = \gamma_i^{\tau} \xi_i^{\tau} + \sum_k{}' \gamma_k^{\tau} \beta_{ik},$$

and thus tell us nothing more about the γ_i^{τ}. In these cases it is most convenient to regard the conjugating system as a series of n localized orbitals, one on each atom, and each containing x electrons.

Relation to the electro-affinity scale

Consider a simple electron pair bond between two univalent atoms a and b, and suppose that the essential features of the preceding analysis are also applicable to such σ bonds. Let us find a physical interpretation of the condition for homopolar bonding between a and b:

$$\gamma_a = \gamma_b, \quad Q_a = Q_b.$$

From equations (19), if $\psi = \gamma_a \phi_a + \gamma_b \phi_b$ is the bonding molecular orbital, we have

$$\gamma_a(\xi_a - \lambda) + \gamma_b \beta_{ab} = 0, \quad \gamma_b(\xi_b - \lambda) + \gamma_a \beta_{ab} = 0,$$

where $$\xi_i = \eta_i(\gamma_i^2) = -I_i + (I_i - E_i)\gamma_i^2 \quad (i = a, b).$$

Eliminating λ, it is found that

$$\gamma_a \gamma_b (\xi_a - \xi_b) = \beta_{ab}(\gamma_a^2 - \gamma_b^2).$$

In order that $\gamma_a = 1/\sqrt{2} = \gamma_b$, it is now clear that we must have

$$-I_a + \tfrac{1}{2}(I_a - E_a) = -I_b + \tfrac{1}{2}(I_b - E_b)$$

or $$I_a + E_a = I_b + E_b. \tag{23}$$

Thus our criterion for equal electro-affinities of atoms a and b, or for homopolar bonding between a and b, is that the sum of the (valence state) ionization potential and the electron affinity should be the same for both atoms. If the condition (*d*) defining the atomic fields v_i^{τ} be dropped, and the overlap integral—which is large for σ bonds—be included in the argument, the result (23) may again be derived from the assumption (12). This criterion is identical with that of Mulliken (1934*b*), who derived it with certain plausible simplifying assumptions by means of the pairing approximation. Thus he showed that, if Ψ_{ab} is the antisymmetrical HLSP function for the electron pair bond between a and b, and $\Psi(a-b)$, $\Psi(a^+b^-)$, $\Psi(a^-b^+)$ are the corresponding covalent and two-ionic structures respectively, then the condition that $\alpha = \beta$ in

$$\Psi_{ab} = \Psi(a-b) + \beta\Psi(a^+b^-) + \alpha\Psi(a^-b^+)$$

is also $I_a + E_a = I_b + E_b$. The absolute electro-affinity scale which may be set up using $\tfrac{1}{2}(I + E)$ has been shown by Mulliken to correspond closely, in those cases where the comparison may be made, to that of Pauling. A discussion of the LCAO molecular orbital approximation, including an approximate derivation of Pauling's thermochemical scale, has also been given by Mulliken in the light of this. In this paper (1935) a procedure for estimating charge distribution was outlined, but this was only semi-quantitative and neither very general nor very satisfactory.

DISCUSSION AND COMPARISON WITH THE SECULAR EQUATION METHOD

It has been customary hitherto (Wheland & Pauling 1935; Lennard-Jones 1937) to use the same fixed set of Coulomb integrals, $\epsilon_i^\tau = \epsilon_i$ (all τ) for each atom i, for all the molecular orbitals of a given configuration. That is, each molecular orbital ψ^τ has been expanded in terms of the same set of atomic functions ϕ_i. The variational theorem, in this case, leads to the familiar secular equations

$$\gamma_i^\tau\left(\epsilon_i - \frac{1}{n^\tau}\lambda^{\tau\tau}\right) + \sum_j{}' \gamma_j^\tau \beta_{ij} = 0 \quad (j \neq i), \tag{24}$$

where $\frac{1}{n^\tau}\lambda^{\tau\tau}$ is the energy of the molecular orbital ψ^τ. It is perhaps not generally realized that the same set of equations is obtained both by minimizing the energy of each individual molecular orbital separately and by minimizing the total zeroth-order energy of the molecular configuration. (It is a property of this method that since $\epsilon_i^\tau = \epsilon_i$ for all τ, we may always arrange for the matrix $[\lambda^{\tau\rho}]$ to be diagonal.) These secular equations may be compared with the equations (20), to which our treatment may be reduced in the absence of orthogonality complications. Whereas the set (24) is linear, and leads to one and the same determinantal equation, among whose roots are all the $\frac{1}{n^\tau}\lambda^{\tau\tau}$, our equations are in general of the fifth degree and admit of no such simplification. What may be called the secular equation method has recently been generalized and extended in application by Coulson & Longuet-Higgins (1947). Unfortunately, however, the relation of the Coulomb integrals ϵ_i to observable quantities is not at all well defined, and, as these authors remarked, it is a pity that no precise interpretation can be given to them.

If the ϵ_i are to be characteristic only of the π orbital of atom i, irrespective of the nature of the conjugated system in which this atom participates, then the method has the severe restriction that the zeroth-order energy of each individual molecular orbital is minimized without reference to the total energy of the π electron configuration. On the other hand, if it is acknowledged that a different set of ϵ_i must be used for each molecular state, then much of the utility of the method must be abandoned, for the choice of the ϵ_i will in general be different for each configuration—so that the comparison of different configurations will be complicated considerably.

In any zeroth-order molecular orbital theory of this type, exchange terms are necessarily only included by subsequent antisymmetrization (cf. Wheland 1938). But there are good grounds for believing that many of the physical properties of molecules emerge without this complication. In this case, therefore, the mutual interactions of the various π-bonding electrons are represented in the zeroth-order treatment by the interdependence of their self-consistent fields. The secular equation method is, however, insufficiently elastic to permit the inclusion of these Hartree interactions. Self-consistency in a rather restricted sense has been demonstrated by Coulson & Rushbrooke (1940) for a large class of aromatic hydrocarbons; it was shown that the assumption of equal ϵ_i's for these molecules in their ground states, and in certain of their excited states, leads to a uniform resultant charge distribution.

But, of course, even for such conjugated hydrocarbons, our treatment does not in general reduce to the secular equation method. In cases of very high symmetry, e.g. benzene and graphite, where the wave functions may be completely specified by group theory, both methods will predict the same (uniform) charge distribution and bond orders. For molecules of lower symmetry, however, this will no longer be true. Consider, for example, butadiene; as is customary, we assume that the σ electron distribution is uniform and may be neglected here. The π electron configuration for the ground state may be written

$$[\psi^{\epsilon}]^2\,[\psi^{\iota}]^2,$$

where ψ^{ϵ}, ψ^{ι} transform as irreducible representations B_1, A_2 or A_u, B_g respectively, according as we are considering the *cis* (C_{2v}) or *trans* (C_{2h}) form of butadiene. Indexing the carbon atom (a, b, c, d) in that order as we pass from one end of the chain to the other, the LCAO forms are

$$\psi^{\epsilon} = \gamma^{\epsilon}_a\phi^{\epsilon}_a + \ldots + \gamma^{\epsilon}_d\,\phi^{\epsilon}_d, \quad \psi^{\iota} = \gamma^{\iota}_a\phi^{\iota} + \ldots + \gamma^{\iota}_d\,\phi^{\iota}_d,$$

where by symmetry and normalization we have

$$2(\gamma^{\epsilon}_a)^2 + 2(\gamma^{\epsilon}_b)^2 = 1 = 2(\gamma^{\iota}_a)^2 + 2(\gamma^{\iota}_b)^2.$$

Equations (19) become

$$\gamma^{\epsilon}_a\xi^{\epsilon}_a + \gamma^{\epsilon}_b\beta - \lambda^{\epsilon\epsilon}\gamma^{\epsilon}_a = 0, \qquad \gamma^{\iota}_a\xi^{\iota}_a + \gamma^{\iota}_b\beta - \lambda^{\iota\iota}\gamma^{\iota}_a = 0,$$

$$\gamma^{\epsilon}_b\xi^{\epsilon}_b + (\gamma^{\epsilon}_a + \gamma^{\epsilon}_b)\,\beta - \lambda^{\epsilon\epsilon}\gamma^{\epsilon}_b = 0, \quad \gamma^{\iota}_b\xi^{\iota}_b + (\gamma^{\iota}_a - \gamma^{\iota}_b)\,\beta - \lambda^{\iota\iota}\gamma^{\iota}_b = 0,$$

where $$\xi^{\tau}_i = -I + (I - E)\,\{q_i - (\gamma^{\tau}_i)^2\} \quad (\tau = \epsilon, \iota;\ i = a, b).$$

Eliminating $\lambda^{\epsilon\epsilon}$, $\lambda^{\iota\iota}$ and setting $\sqrt{2}\,\gamma^{\epsilon}_a = \sin\phi$, $\sqrt{2}\,\gamma^{\iota}_a = \cos\theta$, we find

$$(I - E)\,(2\cos 2\theta - \cos 2\phi) - 2\beta + 4\beta\cot 2\phi = 0,$$

$$(I - E)\,(\cos 2\theta - 2\cos 2\phi) + 2\beta - 4\beta\cot 2\theta = 0.$$

On adding these equations, it is readily verified that, since we are only interested in solutions satisfying the inequality $0 \leqslant \phi$, $\theta \leqslant \frac{1}{4}\pi$ and β is negative, so $\theta = \phi$. Accordingly our equations reduce to

$$(I - E)\cos 2\theta - 2\beta + 4\beta\cot 2\theta = 0,$$

which possesses only one real root in the required interval; the numerical value of this root will depend on the ratio $(I - E)/\beta$. The charge distribution and bond orders are given by

$$q_a = 2(\gamma^{\epsilon}_a)^2 + 2(\gamma^{\iota}_a)^2 = \sin^2\phi + \cos^2\theta = 1,$$

$$q_b = 2(\gamma^{\epsilon}_b)^2 + 2(\gamma^{\iota}_b)^2 = \cos^2\phi + \sin^2\theta = 1,$$

$$p_{ab} = 2(\gamma^{\epsilon}\gamma^{\epsilon}_b + \gamma^{\iota}_a\gamma^{\iota}_b) = \sin\phi\,\cos\phi + \sin\theta\,\cos\theta = \sin 2\theta,$$

$$p_{bc} = 2(\gamma^{\epsilon}_b\gamma^{\epsilon}_b - \gamma^{\iota}_b\gamma^{\iota}_b) = \cos^2\phi - \sin^2\theta = \cos 2\theta.$$

According to the secular equation method, the bond orders are independent of the choice of the (negative) resonance integral β (Coulson 1939). The same bond orders for butadiene are predicted by both methods only if $|\beta| \gg |(I-E)|$. Since, however, $(I-E) \approx 10\,\mathrm{eV}$, the bond order of the central link is increased at the expense of that of the outer links. It is interesting that both methods agree in assigning a uniform charge distribution to this particular molecule.

The author would like to thank Professor Coulson for helpful advice and the Board of the British Rubber Producers' Research Association for a research grant.

REFERENCES

Coulson, C. A. 1939 *Proc. Roy. Soc.* A, **169**, 413.
Coulson, C. A. & Rushbrooke, G. S. 1940 *Proc. Camb. Phil. Soc.* **36**, 193.
Coulson, C. A. & Longuet-Higgins, H. C. 1947 *Proc. Roy. Soc.* A, **191**, 39.
Goeppert-Mayer, M. & Sklar, A. 1938 *J. Chem. Phys.* **6**, 645.
Hartree, D. R. 1928 *Proc. Camb. Phil. Soc.* **24**, 89, 426.
Hund, F. 1933 *Handbuch der Physik*, xxiv/1, 561.
Lennard-Jones, J. E. 1937 *Proc. Roy. Soc.* A, **138**, 280.
Moffitt, W. 1949 *Proc. Roy. Soc.* A, **196**, 524.
Mulliken, R. S. 1934*a* *Phys. Rev.* **46**, 549.
Mulliken, R. S. 1934*b* *J. Chem. Phys.* **2**, 782.
Mulliken, R. S. 1935 *J. Chem. Phys.* **3**, 573.
Mulliken, R. S. 1939 *J. Chem. Phys.* **7**, 14, 20, 121, 339.
Slater, J. C. 1929 *Phys. Rev.* **34**, 1293.
Slater, J. C. 1930 *Phys. Rev.* **36**, 56.
Van Vleck, J. H. 1934*a* *J. Chem. Phys.* **2**, 20, 297.
Van Vleck, J. H. 1934*b* *Phys. Rev.* **45**, 405.
Wheland, G. W. 1938 *Proc. Roy. Soc.* A, **164**, 397.
Wheland, G. W. & Pauling, L. 1935 *J. Amer. Chem. Soc.* **57**, 2091.

The electronic structures of carbon monoxide and carbon dioxide

By W. Moffitt, *New College, Oxford*

(*Communicated by F. E. Simon, F.R.S.—Received* 6 *August* 1948)

By means of the self-consistent LCAO method, electronic structures are assigned to the ground states of carbon monoxide and carbon dioxide, to their ions and to their more important excited states (CO: $X\,^1\Sigma^+$, $A\,^1\Pi$, $B\,^1\Sigma^+$; CO$^+$: $X\,^2\Sigma^+$, $A\,^2\Pi$, $B\,^2\Sigma^+$; CO_2, $^1\Sigma_g^+$; CO_2: $^2\Pi_g$, $^2\Pi_u$, $^2\Sigma_u^+$). Carbon monoxide, which exhibits many unusually interesting features, has been treated at some length. It has been possible to compare, systematically, the bonding in these different systems and to correlate the various carbon-oxygen internuclear distances. The analysis of the divalent states of carbon has been pursued in detail, since this will also have a bearing on the structure of the isonitriles.

Introduction

It is proposed to examine the molecules CO, CO^+ and CO_2, CO_2^+ in their ground states and in certain of their excited states. The aim of the discussion will be to assign plausible electronic structures to these systems on mainly theoretical grounds.

General considerations

We shall use the molecular orbital method and discuss molecular states as arising from orbital configurations. Unless two or more alternative configurations, which are not widely separated energetically (in zeroth-order approximation, say), each give rise to Ψ_K^R's of the same symmetry, we shall assume that molecular states may be described by means of one molecular orbital configuration only. Otherwise states must be described by the superposition of functions Ψ_K^R from each of the feasible constituent configurations. By Ψ_K^R we mean that many-electron wave function, including spin, which may be constructed from the product functions of a configuration K, so as to have total quantum numbers R; R is a threefold symbol representing, first, an irreducible representation of the molecule's spatial symmetry group; secondly, a representation of the appropriate symmetric group (or, alternatively, since we assume a spin-free Hamiltonian, of the spin-rotation group); and lastly, an ordering symbol.

Since all the systems with which we shall be dealing are linear, individual molecular orbitals and the electrons associated with them may be classified as σ or π according to their transformation properties—or, equivalently, according as they have zero or unit components of angular momentum (in units of $\hbar$) about the internuclear axis. (We shall have no occasion to invoke the contribution of d electrons.) The constituent atomic orbitals of a σ electron pair bond will overlap to a greater extent than those constituting a π bond. We may therefore reasonably suppose that the resonance integrals for σ-bonding electrons are considerably greater, numerically, than those for π-bonding electrons. And from this it is readily shown that any heteropolar character in a multiply-bonded system will tend to be assimilated by its π electrons. Briefly, π electronic shells are more readily polarized than σ electron shells—for which there is, of course, also much empirical justification. In order to obtain quantitative estimates of π bond strength and charge distribution, we shall use the self-consistent LCAO method (Moffitt 1949). This method has been shown to lead to the absolute electro-affinity scale introduced by Mulliken (1934*a*), and we shall therefore use this scale in discussing σ electrons.

When describing polyatomic molecules, such as carbon dioxide, it is sometimes more convenient to use not the molecular orbitals themselves, but so-called 'localized' molecular orbitals (Hund 1933). This procedure implies the molecular orbital equivalent to the principle of additivity—it presupposes that some of the molecule's electronic charge distribution may be divided up into domains (generally electron pair bonds or 'inert' pairs), the properties of each of which are adequately described by an antisymmetric two-electron function. These localized one-electron orbital functions, while representing charge distribution and bond strength, will no longer transform as if they belonged to irreducible representations of the symmetry group of the molecule as a whole. On the contrary, as a set, they may be regarded as forming the basis for a reducible representation of this group. The Hartree energy parameters associated with such functions will not in general, therefore, be directly related to term values of the molecular spectrum.

For the σ electrons of diatomic systems and, if we use localized molecular orbitals, for those of polyatomic systems, the atomic orbitals from which the molecular and localized molecular orbitals are constructed, may be characterized, *inter alia*, by hybridization ratios. These determine the proportions of s and p character to be associated with the various atomic orbitals, and are obtained from symmetry considerations or by application of the variational principle—that is, by treating them as parameters with respect to which the energy is minimized (e.g. Coulson & Moffitt 1949).

The absolute electro-affinity and radius of an atomic orbital are functions of the hybridization ratio associated with this atomic orbital. An estimate of the former property may be made from spectroscopic data on the atom under consideration (Mulliken 1934*a*). It is found that, with increasing s character, the electro-affinity of an atomic orbital increases considerably. Using Hartree field functions or simplified analytical expressions (Slater 1930) for the atomic $2s$ and $2p$ orbitals of carbon, it has been shown that the radius of a hybrid orbital increases with increasing p character (Moffitt & Coulson 1947, and, more particularly, Coulson 1948). Elementary screening considerations suggest that, since Slater functions contain the same radial distribution for both s and p orbitals, this effect may be even more pronounced than Coulson's calculations would suggest. At short distances, therefore, we should expect the s atomic orbitals to be more strongly σ-bonding than the p orbitals. Consider a σ-bonding electron pair which is strained in the sense that other factors, such as high π-bonding strength, have reduced the internuclear separation to some value below its equilibrium single σ bond distance. Then, *ceteris paribus* and providing the hybridization ratios of the σ-type atomic orbitals have not been preemptively selected by valence or symmetry considerations, there will be a tendency to increase the s character of the bond—even at the expense of some promotional energy which will, in general, be required to produce the appropriate atomic valence states.

Lastly it should be explained in what sense we shall apply the terms bonding, non-bonding or antibonding to molecular orbitals. If the contribution from some orbital to the total bond order between two atoms is positive, zero or negative, then we shall use these terms respectively to describe the behaviour of such an orbital between these atoms. Our usage corresponds to the usual analytical definition, but does not coincide with the customary classification of orbitals according to their spectroscopic properties. Thus the removal of a non-bonding electron from a diatomic molecule will, in general, lead to a strengthening or a weakening of the bonding between the two atoms owing to the reorganization of the residual electronic configuration.

CONCERNING DIVALENT CARBON

In its ground state, ^{3}P, the carbon atom in free space may be described as arising from the configuration $(1s)^2\,(2s)^2\,(2p)^2$, the orientation of the spins of the $2p$ electrons being parallel. The lowest divalent state ($\overline{V}_2^{\dagger}$) is obtained by uncoupling the spins of these two electrons, so that their mutual inclination is random (with $\overline{\boldsymbol{\sigma}_i . \boldsymbol{\sigma}_j} = 0$ in the notation of Dirac 1935). Another divalent state ($\overline{V}_2^{\S}$) may be similarly constructed from the lowest state, ^{5}S, of carbon with configuration $(1s)^2\,(2s)\,(2p)^3$. The quadrivalent states which arise have been discussed by Van Vleck (1934).

Now let us suppose the carbon atom to be placed in a field which has rotational symmetry ($C_{\infty v}$) about the z-axis. If this field is weak, we may continue to discuss the configurations of these valence states in terms of one-electron $1s$, $2s$ and $2p$ orbitals—that is, the angular momentum, l, associated with each of these orbitals remains a good quantum number. The $2p$ degeneracy is partially removed, and we may refer to a π or to a σ electron, according as the component of its angular momentum about the z-axis is unity ($m_l = \pm 1$) or vanishes ($m_l = 0$). We shall be most interested in those states of the divalent carbon atom whose two valence electrons are respectively σ and π, and whose remaining electrons have σ symmetry. Thus from $\overline{V}_2^{\dagger}$, we obtain the valence state $V_2^{\dagger}$: $(1s)^2\,(2s)^2\,(2p\sigma^*)\,(2p\pi^*)$, the asterisk (*) denoting a valence orbital; similarly from $\overline{V}_2^{\S}$ we obtain $V_2^{\S}$: $(1s)^2\,(2s^*)\,(2p\sigma)^2\,(2p\pi^*)$.

Now let the field strength be increased. In general, since $\overline{V}_2^{\dagger}$, $\overline{V}_2^{\S}$ are the only low-lying divalent states of carbon, we may treat the lowest resulting valence state V_2 as a mixture of $V_2^{\dagger}$ and $V_2^{\S}$. V_2 may no longer be regarded as arising from either of the configurations $(1s)^2\,(2s)^2\,(2p)^2$ or $(1s)^2\,(2s)\,(2p)^3$, since the distinction between $2s$ and $2p$ electrons has been destroyed. Retaining the atomic orbital approximation, this divalent state may be most conveniently treated as V_2: $(1s)^2\,(2u\sigma)^2\,(2t\sigma^*)\,(2p\pi^*)$, where $2u\sigma$, $2t\sigma$ are mutually orthogonal hybrid $2s$-$2p$ orbitals. We have supposed that the field is not sufficiently strong to affect the K shell electrons, which we shall therefore omit from subsequent discussion. We may write

$$\phi(2t\sigma) = \cos\chi\,\phi(2s) + \sin\chi\,\phi(2p\sigma), \quad \phi(2u\sigma) = \sin\chi\,\phi(2s) - \cos\chi\,\phi(2p\sigma),$$

where $0 \leqslant \chi \leqslant \frac{1}{2}\pi$. Symbolically, therefore,

$$V_2 = \sin^2\chi\, V_2^{\dagger} + \cos^2\chi\, V_2^{\S} = (1-x)\, V_2^{\dagger} + xV_2^{\S},$$

arises from the configuration $(2s)^{2-x}\,(2p)^{2+x}$; when $x = 0$, $V_2 = V_2^{\dagger}$ and when $x = 1$, $V_2 = V_2^{\S}$. The best value χ, in any particular case, will depend upon the details of the field, treated as a perturbation. From spectroscopic data on the carbon atom, it may be shown (Mulliken 1934*a*) that the energies of $V_2^{\dagger}$, $V_2^{\S}$ in very weak fields are respectively some 0·5 and 9·9 eV above that of the ground ^{3}P state (^{5}S probably lies about 4·3 eV above ^{3}P). For intermediate values of x, therefore, we should expect the purely intra-atomic energy of V_2 to lie somewhere between these two values: say, rather less than $0{\cdot}5(1-x) + 9{\cdot}9x$ eV as a first estimate.

It will be useful, in discussing the structure of carbon monoxide and other molecules, to have values for the energies associated with the three processes

$$\mathrm{C}, V_2 \xrightarrow{E^\circ} \mathrm{C}^-, V_1\!: \quad (2u\sigma)^2\,(2t\sigma)^2\,(2p\pi^*),$$

$$\mathrm{C}, V_2 \xrightarrow{I^\circ} \mathrm{C}^+, V_1\!: \quad (2u\sigma)^2\,(2p\pi^*),$$

$$\mathrm{C}, V_2 \xrightarrow{I^*} \mathrm{C}^+, V_3\!: \quad (2u\sigma^*)\,(2t\sigma^*)\,(2p\pi^*).$$

E°, I° represent the electron affinity and ionization potential respectively of the orbital $(2t\sigma)$, and Mulliken's (1934*a*) tables suggest the estimates (in eV)

$$E^\circ \sim 8{\cdot}7\cos^2\chi, \quad I^\circ \sim 21\cos^2\chi + 10{\cdot}7\sin^2\chi.$$

The absolute electro-affinity associated with this orbital is therefore

$$\tfrac{1}{2}(I^\circ + E^\circ) \sim 14{\cdot}9\cos^2\chi + 5{\cdot}4\sin^2\chi.$$

Owing to uncertainties in the $2s$ term values, more refined calculations are probably not worth while; for intermediate values of χ, the errors should not exceed 0·5 eV or so. The ionization potential of a non-bonding $(2u\sigma)$ electron from V_2 is similarly given by $I^* \sim 18{\cdot}8\sin^2\chi + 11{\cdot}2\cos^2\chi$.

It will have been noticed that, by restricting χ to the interval $(0, \frac{1}{2}\pi)$, we have ensured that the hybrid valence orbital $(2t\sigma^*)$ points in the positive direction of the z-axis. On the other hand, the lone pair of $(2u\sigma)$ electrons will be more densely distributed in the opposite direction. And so the valence state V_2 will, in general, have a non-vanishing intrinsic dipole moment which may be called the atomic dipole μ_{C} of C, V_2; it will be a function $\mu_{\mathrm{C}}(\chi)$ of the hybridization parameter χ. To make this more explicit, let us find where, along the z-axis, the centroid of the electronic distribution will lie. Clearly it is only the 2σ electrons whose contribution need be considered, and for these

$$\begin{aligned} \bar{z}(\chi) &= \int \phi(2t\sigma^*)\, z\, \phi(2t\sigma^*)\, d\tau + 2\int \phi(2u\sigma)\, z\, \phi(2u\sigma)\, d\tau \\ &= -\sin 2\chi \int \phi(2s)\, z\, \phi(2p\sigma)\, d\tau. \end{aligned}$$

$\mu_{\mathrm{C}}(\chi) = e\bar{z}(\chi)$ will therefore only vanish when $\chi = 0$ or $\frac{1}{2}\pi$, that is, when V_2 becomes purely $V_2^{\dagger}$ or $V_2^{\S}$. Since the integral $\int \phi(2s)\, z\, \phi(2p\sigma)\, d\tau$ is positive, the atomic dipole will be directed along the negative z-axis whenever $0<\chi<\frac{1}{2}\pi$. Using analytical atomic functions (Zener 1930), the numerical value of this integral may be estimated with reasonable accuracy at 0·5 Å. When $\chi=\frac{1}{4}\pi$, therefore, $|\mu_{\mathrm{C}}|$ attains its maximum value of 2·4 Debye units.

THE STRUCTURE OF CARBON MONOXIDE

The carbon atom in the ground state $X\,^1\Sigma^+$ of carbon monoxide is considered to be in a divalent state V_2. Owing to the high promotional energy required to produce states arising from the atomic configuration $(2s)\,(2p)^5$ for oxygen, we shall suppose that hybridization is only important for the carbon atom. At large internuclear separations, the appropriate carbon V_2 will be $V_2^\dagger$. As this distance is decreased, however, the σ-bonding strength will be enhanced by decreasing, from $\frac{1}{2}\pi$, the hybridization parameter χ of $V_2=\sin^2\chi\,V_2^\dagger+\cos^2\chi\,V_2^\S$.

The qualitative features of the ultra-violet spectrum of CO, $X\,^1\Sigma^+$ have been successfully explained (Mulliken 1932) in terms of the configuration $\sigma^2\sigma^2\pi^4\sigma^2$, or, more completely,

$$\mathrm{CO},\ X\,^1\Sigma^+:\quad (K)\,(K)\,(s\sigma)^2\,(t\sigma)^2\,(\omega\pi)^4\,(u\sigma)^2.$$

We shall describe these molecular orbitals by assigning plausible LCAO forms to them. Since the carbon atom is to be divalent and the oxygen atom is to be in a valence state derived from the configuration $(2s)^2\,(2p)^4$, two of the σ electron pairs must correspond to an 'inert pair' of carbon electrons and a pair of oxygen $2s$ electrons. Accordingly we write

$$\psi(s\sigma)=\phi_{\mathrm{O}}(2s)\quad\text{and}\quad\psi(u\sigma)=\phi_{\mathrm{C}}(2u\sigma).$$

The LCAO forms for the remaining molecular orbitals must therefore be

$$\psi(t\sigma)=\gamma^t_{\mathrm{O}}\phi_{\mathrm{O}}(2p\sigma)+\gamma^t_{\mathrm{C}}\phi_{\mathrm{C}}(2t\sigma),\quad\psi(\omega\pi)=\gamma^\omega_{\mathrm{O}}\phi_{\mathrm{O}}(2p\pi)+\gamma^\omega_{\mathrm{C}}\phi_{\mathrm{C}}(2p\pi),$$

where the coefficient- and hybridization-parameters, γ^τ_i ($\tau=t,\,\omega;\,i=\mathrm{O},\mathrm{C}$) and χ, are still to be determined. (More strictly, since the three σ molecular orbitals are not orthogonal, we should have to construct a set of three slightly different and mutually orthogonal orbitals from these; if we were to attempt accurate calculations of the σ electron energy, such a process would be essential. Since we are immediately only concerned with the total σ electron distribution, it is reasonable, however, to suppose that we may arrive at reliable semi-quantitative conclusions without this refinement.)

Consider, first, the hybridization parameter χ. At relatively large internuclear separations ($r>1{\cdot}5$ Å, say), hybridization will be unimportant and $\phi_{\mathrm{C}}(2t\sigma)$ will be almost entirely $\phi_{\mathrm{C}}(2p\sigma)$. As r approaches its equilibrium value of $r_e=1{\cdot}1284$ Å, however, a compromise between strong σ-type bonding and the energy of promotion required to produce the appropriate C, V_2 sets in. We shall suppose that in the neighbourhood of r_e, χ is changing fairly rapidly and has a value $\chi\sim\sin^{-1}\frac{1}{2}\sqrt{3}$ at $r=r_e$. Our particular value gives the *a priori* expected order for the hybridization and is consistent (*vide infra*) with the molecule's chemical properties and electronic charge distribution. In this case, $\phi_{\mathrm{C}}(2t\sigma)$ will resemble closely the carbon atom orbitals used in constructing localized molecular orbitals for methane. The carbon atom in CO, $X\,{}^1\Sigma^+$, may therefore be considered as arising from the atomic configuration $(2s)^{1\frac{3}{4}}\,(2p)^{2\frac{1}{4}}$. This places an upper limit of 17 eV on the 'ionization potential' of $\phi_{\rm C}(2u\sigma)$ and gives an absolute electro-affinity of ~ 8 eV for the $\phi_{\rm C}(2t\sigma)$ orbital.

Next we consider the values of the coefficients $\gamma^t_{\rm O}$, $\gamma^t_{\rm C}$. The absolute electro-affinities of oxygen's $\phi_{\rm O}(2p\sigma)$ and $\phi_{\rm O}(2p\pi)$ orbitals are some 10·0 and 9·4 eV respectively. We have seen that in general π electrons are more readily polarized than σ electrons. Since, in addition, the electro-affinities of $\phi_{\rm O}(2p\sigma)$ and $\phi_{\rm C}(2t\sigma)$ only differ by 2 eV, whereas those of $\phi_{\rm O}(2p\pi)$ and $\phi_{\rm C}(2p\pi)$ differ by $9{\cdot}4-5{\cdot}4 = 4$ eV, it is reasonable to suppose that the four π electrons account for nearly all the heteropolar character of the bonding, and that the electron pair bond $(t\sigma)^2$ is largely homopolar with $\gamma^t_{\rm O} \approx \gamma^t_{\rm C}$. A good estimate of the σ electron dipole moment is therefore simply $\mu(\sigma) = \mu_{\rm C} = -2{\cdot}4\sin 2\chi \approx -2{\cdot}1$ Debye units; a more detailed calculation would encounter orthogonalization difficulties, the attack of which our approximation scarcely warrants. This contribution of the σ electrons is directed from the oxygen atom towards the carbon atom and may be attributed to the hybridization at the carbon atom.

Lastly, let us deal with the $(\omega\pi)$ electrons. The parameters $\gamma^{\omega}_{\rm C}$, $\gamma^{\omega}_{\rm O}$ may be determined by means of the 'self-consistent' LCAO method which has been described recently (Moffitt 1949). The best choice of these then satisfies the equations

$$\gamma^{\omega}_{\rm C}\xi^{\omega}_{\rm C}+\gamma^{\omega}_{\rm O}\beta_{\rm CO}-\lambda\gamma^{\omega}_{\rm C} = 0, \quad \gamma^{\omega}_{\rm O}\xi^{\omega}_{\rm O}+\gamma^{\omega}_{\rm C}\beta_{\rm CO}-\lambda\gamma^{\omega}_{\rm O} = 0.$$

On setting $\gamma^{\omega}_{\rm C} = \sin\theta$ and eliminating λ, we find that $\gamma^{\omega}_{\rm O} = \cos\theta$ and

$$(\xi^{\omega}_{\rm C}-\xi^{\omega}_{\rm O})+2\beta_{\rm CO}\cot 2\theta = 0.$$

Now the four $(\omega\pi)$ electrons are to be fitted into a skeleton consisting of the two atomic nuclei and the σ electrons $(K)\,(K)\,(s\sigma)^2\,(t\sigma)^2\,(u\sigma)^2$; the field due to this system gives a single positive electronic charge on the carbon atom and three positive charges on the oxygen atom. Accordingly $\xi^{\omega}_{\rm C}$, $\xi^{\omega}_{\rm O}$, which determine the self-consistent Coulomb terms, are given by

$$\xi^{\omega}_{\rm C} = -I_{\rm C}+(I_{\rm C}-E_{\rm C})\,\{3(\gamma^{\omega}_{\rm O})^2\} = -I_{\rm C}+(I_{\rm C}-E_{\rm C})\,(3\sin^2\theta),$$

$$\xi^{\omega}_{\rm O} = -I_{\rm O}+(I_{\rm O}-E_{\rm O})\,\{3(\gamma^{\omega}_{\rm O})^2-2\} = -I_{\rm O}+(I_{\rm O}-E_{\rm O})\,(3\cos^2\theta-2),$$

where $I_{\rm C}$, $I_{\rm O}$ and $E_{\rm C}$, $E_{\rm O}$ are the ionization potentials and electron affinities of $\phi_{\rm C}(2p\pi)$ and $\phi_{\rm O}(2p\pi)$ respectively. Numerical estimates (in eV) of the quantities may be formed from spectroscopic data on the carbon and oxygen atoms, from which it is found that

$$I_{\rm C} = 11{\cdot}0, \quad I_{\rm O} = 15{\cdot}8, \quad E_{\rm C} = 0{\cdot}5, \quad E_{\rm O} = 2{\cdot}9.$$

By consideration of a series of carbon-oxygen bonds, it seems best to take $\beta_{\rm CO} = -3$ eV, a value which is very reasonable, *a priori*. (It is a feature of this self-consistent LCAO method that small variations in the choice of the resonance integrals do not alter the charge distribution and bond orders greatly.) We must therefore solve the equation

$$(I_{\rm C}+3I_{\rm O}-3E_{\rm C}-E_{\rm O})-3\cos 2\theta[(I_{\rm C}-E_{\rm C})+(I_{\rm O}-E_{\rm O})]+4\beta_{\rm CO}\cot 2\theta = 0,$$

for θ in the interval $(0, \frac{1}{2}\pi)$, with these values. The π bond order is then

$$p^{\omega}_{\mathrm{CO}} = 4\gamma^{\omega}_{\mathrm{C}}\gamma^{\omega}_{\mathrm{O}} = 2\sin 2\theta,$$

and the net formal charges associated with the carbon and oxygen atoms respectively are $Q_{\mathrm{C}} = 2\cos 2\theta - 1$ and $Q_{\mathrm{O}} = 1 - 2\cos 2\theta$. The solution we seek is

$$\theta = \tfrac{1}{2}\cos^{-1} 0{\cdot}629,$$

so that $$p^{\pi}_{\mathrm{CO}} = p^{\omega}_{\mathrm{CO}} = 1{\cdot}554, \quad Q_{\mathrm{C}} = 0{\cdot}259, \quad Q_{\mathrm{O}} = -0{\cdot}259.$$

It is interesting to calculate the dipole moment of the π electrons. The centroid of the negative charge for the $(\omega\pi)^4$ shell electrons lies on the z-axis and has coordinate

$$\bar{z} = \int \psi(\omega\pi)\, z\, \psi(\omega\pi)\, d\tau = (\gamma^{\omega}_{\mathrm{O}})^2 r_e + 2(\gamma^{\omega}_{\mathrm{O}}\gamma^{\omega}_{\mathrm{C}}) \int \phi_{\mathrm{O}}(2p\pi)\, z\phi_{\mathrm{C}}(2p\pi)\, d\tau;$$

the origin of co-ordinates is taken at the carbon nucleus, with the oxygen on the positive z-axis. The π electron dipole moment is therefore

$$\mu(\pi) = er_e\{4(\gamma^{\omega}_{\mathrm{O}})^2 - 3\} + 8(\gamma^{\omega}_{\mathrm{C}}\gamma^{\omega}_{\mathrm{O}}) \int \phi_{\mathrm{O}}(2p\pi)\, z\phi_{\mathrm{C}}(2p\pi)\, d\tau.$$

Here the first term represents the heteropolar dipole moment due to the unequal partition of charge between the singly-charged carbon and the triply-charged oxygen atoms. The second term arises from the inequality in the sizes of the $(2p\pi)$ carbon and oxygen atom orbitals; it has been called the homopolar dipole term by Mulliken (1935*a*). The heteropolar moment is easily shown to be directed from the carbon to the oxygen atom, and has a magnitude of ~ 1·3 Debye units. In order to calculate the homopolar moment, we use approximate analytical expressions for the atomic wave functions (Slater 1930), namely,

$$\phi_i(2p\pi) = \mathcal{N}_i r_i \cos\phi_i \sin\theta_i \exp(-m_i r_i) \quad (i = \mathrm{C}, \mathrm{O}),$$

and introduce a correction for their non-orthogonality; the normalizing factors have the values $\mathcal{N}_i = \sqrt{(m_i^5/\pi)}$ and $z = r_{\mathrm{C}}\cos\theta_{\mathrm{C}}$. The homopolar integral may be evaluated in closed form by transforming to spheroidal co-ordinates λ, μ, ϕ, where

$$\lambda = (r_{\mathrm{C}} + r_{\mathrm{O}})/r_e \quad (1 \leqslant \lambda < \infty), \qquad \mu = (r_{\mathrm{C}} - r_{\mathrm{O}})/r_e \quad (-1 \leqslant \mu \leqslant +1),$$

and $\phi_{\mathrm{C}} = \phi_{\mathrm{O}} = \phi$ $(0 \leqslant \phi \leqslant 2\pi)$, the aximuth around the z-axis. The integration reduces to the evaluation of the quantities

$$A_n(\alpha) = \int_1^{\infty} \lambda^n e^{-\alpha\lambda}\, d\lambda, \quad \alpha = \tfrac{1}{2}(m_{\mathrm{C}} + m_{\mathrm{O}})\, r_e,$$

$$B_n(\beta) = \int_{-1}^{+1} \mu^n e^{-\beta\mu}\, d\mu, \quad \beta = \tfrac{1}{2}(m_{\mathrm{C}} - m_{\mathrm{O}})\, r_e,$$

which is most easily effected for the various values of $n = 0, 1, \ldots, 5$ by means of recursion formulae. (This particular integral is one of the very few which have not been listed by Coulson (1942).) By these means we find, for our homopolar moment, some 0·6 Debye unit in the same direction as the heteropolar moment. The total π dipole moment is therefore $\mu(\pi) = 1{\cdot}3 + 0{\cdot}6 = 1{\cdot}9$ Debye units. (We have not considered induced moments because our self-consistent LCAO method includes these automatically as far as an atomic orbital approximation permits.)

We are now in a position to discuss the dipole moment of the molecule as a whole. At the equilibrium internuclear separation r_e, the calculated dipole moment is

$$\mu_{\mathrm{CO}} = \mu(\sigma)+\mu(\pi) = -2{\cdot}1+1{\cdot}9 = -0{\cdot}2\ \text{Debye unit.}$$

Our small value is in excellent agreement with the observed moment of 0·118 Debye unit and, in addition, gives a direction to this moment. Since the theoretical value depends not unappreciably on the choice of the hybridization parameter χ, a quantity whose value is, as yet, not amenable to direct calculation, we are hardly in a position to insist on the significance of this sign. Nevertheless, the behaviour of the molecule in other respects (*vide infra*) does suggest that the $(u\sigma)^2$ lone pair has rather more directional properties than would be the case if this sign were positive.

Let us consider the variation of μ_{CO} with the internuclear distance r. Qualitatively, we should expect the dipole moment at about 1·5 Å, say, to be due, almost entirely, to the π electrons alone, $(u\sigma)^2$ becoming essentially $(2s_{\mathrm{C}})^2$; this corresponds to a modified Long & Walsh structure for carbon monoxide (Long & Walsh 1947; Moffitt 1948). As r decreases from 1·5 Å, the amount of carbon $(2s)$ character in $(t\sigma)^2$ increases as does the $(2p)$ character of $(u\sigma)^2$. That is, χ is expected to increase fairly rapidly with r for values of r near r_e. In the neighbourhood of r_e, we shall have, in our approximation,

$$\mu_{\mathrm{CO}} = 1{\cdot}9\left(\frac{r}{r_e}\right)-2{\cdot}4\sin 2\chi,$$

so that

$$r_e\frac{\partial\mu_{\mathrm{CO}}}{\partial r} = 1{\cdot}9-4{\cdot}8r_e\cos 2\chi\frac{\partial\chi}{\partial r} = 1{\cdot}9+2{\cdot}7\frac{\partial\chi}{\partial r}.$$

Now, from an examination of the intensities in the infra-red spectrum of carbon monoxide (Matheson 1932; Mulliken 1934*b*), it may be shown, though not unambiguously, that most probably $r_e(\partial\mu_{\mathrm{CO}}/\partial r) = 9{\cdot}1$. Accordingly, in order that our model correspond to the observed variations, we must take $\partial\chi/\partial r = 2{\cdot}7$; this confirms our qualitative predictions. The large spectral intensities may therefore be consistently explained in terms of the varying σ electron charge distribution due to changes of hybridization. It is difficult to see how displacements in the 'Pauling equilibrium' (1932) $\{\overset{-}{\mathrm{C}}\overset{\leftarrow}{\equiv}\overset{+}{\mathrm{O}}, \mathrm{C}{=}\mathrm{O}, \overset{+}{\mathrm{C}}{-}\overset{-}{\mathrm{O}}\}$ can account for variations of this order.

The chemical reactivity of carbon monoxide may be discussed on the basis of this structure. It is well known that carbon monoxide is readily oxidized to carbon dioxide, and that it has electron donor properties which are similar to, but rather weaker than those of ammonia. Thus carbon monoxide enables boron to complete its valence shell in forming $BH_3.CO$, forms carbonyls with metals and complexes with their ions; it also combines with haemoglobin. Since the isonitriles possess similar properties, we may suppose that these reactions are characteristic of divalent carbon. As we have seen, a 'lone pair' of $(2u\sigma)$ electrons is associated with the carbon atom, which gives rise to the large atomic dipole moment, μ_{C}; the ammonia molecule possesses a similar pair of electrons which avoid the H_3 region. The $(u\sigma)^2$ shell is therefore nicely suited, sterically, for combining with groups which approach the molecule, thereby increasing the valence number of the carbon atom. The rearrangement of the CO system accompanying such reactions (compare CO^+, $X\,{}^2\Sigma^+$, below) will assist the progress of their transition states over the potential barrier and reduce what might have been a considerable energy of activation.

THE EXCITED STATE $A\,{}^1\Pi$ OF CARBON MONOXIDE

Consider the upper state $A\,{}^1\Pi$ of the fourth positive bands ($\sim 64{,}500\ \text{cm.}^{-1}$) in the ultra-violet spectrum of carbon monoxide. We shall discuss the structure of this excited state on the basis of the configuration (Sponer 1935)

$$\text{CO},\ A\,{}^1\Pi\text{:}\quad (K)\,(K)\,(s\sigma)^2\,(t\sigma)^2\,(\omega\pi)^4\,(u\sigma)\,(v\pi);$$

it lies some 8 eV above the lower, ground state. As before

$$\psi(s\sigma) = \phi_O(2s) \quad \text{and} \quad \psi(u\sigma) = \phi_C(2u\sigma),$$

but the hybridization parameter χ, which defines $\phi_C(2u\sigma)$ and $\phi_C(2t\sigma)$, will be different to that used in discussing the $X\,{}^1\Sigma^+$ state. The LCAO forms for the remaining orbitals are

$$\begin{aligned}\psi(t\sigma) &= \gamma^t_O\phi_O(2p)+\gamma^t_C\phi_C(2t\sigma),\\ \psi(\omega\pi) &= \gamma^\omega_O\phi^\omega_O(2p\pi)+\gamma^\omega_C\phi^\omega_C(2p\pi),\\ \psi(v\pi) &= \gamma^v_O\phi^v_O(2p\pi)+\gamma^v_C\phi^v_C(2p\pi);\end{aligned}$$

$\psi(\omega\pi)$ is bonding, whereas $\psi(v\pi)$ is antibonding, so that $\gamma^v_O = -\gamma^\omega_C$ and $\gamma^v_C = \gamma^\omega_O$. As before, we shall suppose, with possibly rather less justification, that $\gamma^t_O \approx \gamma^t_C$. The determination of the γ^τ_i ($\tau = \omega, v$; $i = \text{C}, \text{O}$) has been discussed elsewhere (Moffitt 1949). It is found that the π bond order and charge distribution are given by

$$p^\pi_{CO} = 1{\cdot}132, \quad Q_C = +0{\cdot}484, \quad Q_O = -0{\cdot}484.$$

The π bond strength is thus considerably weaker than that in $X\,{}^1\Sigma^+$. This suggests that $\chi > \sin^{-1}\frac{1}{2}\sqrt{3}$, or that less hybridization occurs here. We shall return to discuss the bond length and force constants of this excited state.

THE EXCITED STATE $B\,{}^1\Sigma^+$ OF CARBON MONOXIDE

There remains another excited state, namely, the upper state $B\,{}^1\Sigma^+$ of the Ångström bands, whose structure has not been discussed hitherto, to which we can, however, assign a single plausible configuration. (Clearly our method, which seeks LCAO approximations to the Hartree functions, will be inadequate for dealing with the triplet states of carbon monoxide—for these, antisymmetrization is of paramount importance.) We write

$$\text{CO},\ B\,{}^1\Sigma^+\text{:}\quad (K)\,(K)\,(s\sigma)^2\,(t\sigma)^2\,(\omega\pi)^4\,(u\sigma)\,(r\sigma);$$

$(s\sigma)$, $(t\sigma)$, $(u\sigma)$ will have much the same significance as before; we again assume that $\gamma^t_O \approx \gamma^t_C$ and expect χ to have a value near to that for $X\,{}^1\Sigma^+$. One of the non-bonding $(u\sigma)$ electrons has been excited to the Rydberg orbital $(r\sigma)$, which is larger than the rest of the molecule. We may therefore discuss the structure of this state on the basis of its CO^+ core, which will be very closely related to $X\,{}^2\Sigma^+$, CO^+ (*vide infra*). We thus expect a bond order which is rather less than that of this ion; that is, $p^{\pi}_{CO} \lesssim 1{\cdot}838$. Such a structure enables us to explain the bond shortening which is observed in passing from $X\,{}^1\Sigma^+$ (1·1284 Å) to $B\,{}^1\Sigma^+$ (1·118 Å).

THE GROUND STATE $X\,{}^2\Sigma^+$ OF THE ION CO^+

The first ionization potential of carbon monoxide (~ 14 eV) is attributed to the removal of a non-bonding $(u\sigma)$ electron. The rearrangement of the resulting ion, in particular, the increased π-bonding strength, is to account for its low value (the 'term value', or Hartree energy parameter, associated with $(u\sigma)$ was estimated as $\lesssim 17$ eV). The structure of $X\,{}^2\Sigma^+$ is therefore formulated in terms of the configuration

$$CO^+,\ X\,{}^2\Sigma^+:\quad (K)\,(K)\,(s\sigma)^2\,(t\sigma)^2\,(\omega\pi)^4\,(u\sigma).$$

Such a configuration was put forward by Mulliken (1932); but, in order to account for the increased bond strength in $X\,{}^2\Sigma^+$ (1·115 Å) he assumed that $(u\sigma)$ was associated with antibonding characteristics. We suppose that $(s\sigma)$, $(t\sigma)$ and $(u\sigma)$ are much the same as for CO, $X\,{}^1\Sigma^+$.

Let us determine the π bond order and charge distribution by minimizing the zeroth-order energy with respect to the LCAO coefficients γ^{ω}_{C}, γ^{ω}_{O}. Since the $(\omega\pi)$ electrons are to be introduced into a σ core, with doubly-charged carbon and triply-charged oxygen atoms, we have, on setting $\gamma^{\omega}_{C} = \sin\theta$ $(0 \leqslant \theta \leqslant \frac{1}{2}\pi)$,

$$(-I_C + 3I_O - E_C - E_O) - 3[(I_C - E_C) + (I_O - E_O)]\cos 2\theta + 4\beta_{CO}\cot 2\theta = 0.$$

By solving this equation for θ we find that

$$p^{\pi}_{CO} = 2\sin 2\theta = 1{\cdot}838, \quad Q_C = 2\cos 2\theta = +0{\cdot}788, \quad Q_O = 1 - 2\cos 2\theta = +0{\cdot}212.$$

We notice that $p^{\pi}_{CO}({}^{2}_{X}\Sigma^+, CO^+)$ is considerably greater than $p^{\pi}_{CO}(X\,{}^1\Sigma^+, CO)$.

THE EXCITED STATE $A\,{}^2\Pi$ OF THE ION CO^+

We shall identify the upper state of the comet-tail band system, which occurs as the lower state of the Baldet-Johnson system, as did Mulliken (1928), with the state

$$CO^+,\ A\,{}^2\Pi_{\text{inv.}}:\quad (K)\,(K)\,(s\sigma)^2\,(t\sigma)^2\,(\omega\pi)^3\,(u\sigma)^2;$$

in a higher degree of approximation, of course, this doublet state has components ${}^2\Pi_{1\frac{1}{2}}$, ${}^2\Pi_{\frac{1}{2}}$. By a similar analysis to that used above, it is found that, on setting $\gamma^{\omega}_{C} = \sin\theta$ $(0 \leqslant \theta \leqslant \frac{1}{2}\pi)$,

$$(2I_O - E_O - E_C) - [(I_C - E_C) + (I_O - E_O)]\cos 2\theta + 2\beta_{CO}\cot 2\theta = 0,$$

whence the π bond order and net charge distribution are

$$p^{\pi}_{CO} = 3\sin\theta\cos\theta = 0{\cdot}839, \quad Q_C = 1 - 3\sin^2\theta = +0{\cdot}759,$$

$$Q_O = 3 - 3\cos^2\theta = +0{\cdot}241.$$

The increased bond length of $A\,{}^2\Pi$ ($\sim 1{\cdot}24$ Å) is readily understood.

THE EXCITED STATE $B\,^2\Sigma^+$ OF THE ION CO^+

The upper state of the first negative carbon bands and of the Baldet-Johnson system for CO^+, which has not been discussed previously will be described as arising from the configuration

$$CO^+,\ B\,^2\Sigma^+:\quad (K)\,(K)\,(s\sigma)^2\,(t\sigma)\,(\omega\pi)^4\,(u\sigma)^2.$$

That is, one of the σ-bonding electrons of carbon monoxide has been removed. With $\gamma^{\omega}_{C} = \sin\theta$ as usual, we find

$$2(2I_O - E_C - E_O) - 3[(I_C - E_C) + (I_O - E_O)]\cos 2\theta + 4\beta_{CO}\cot 2\theta = 0,$$

so that

$$p^{\pi}_{CO} = 2\sin 2\theta = 1{\cdot}511, \quad Q_C = 2\cos 2\theta - \tfrac{1}{2} = +0{\cdot}811,$$

$$Q_O = 1\tfrac{1}{2} - 2\cos 2\theta = +0{\cdot}189.$$

The π bond strength is therefore slightly weaker than that in CO, $X\,^1\Sigma^+$, but the σ-bonding is only half as strong. If, therefore, both σ and π electrons were to contribute equally to the first and second derivatives of the total bond energy with respect to the internuclear distance, we should be able to compare the bond distance and force constant of CO^+, $B\,^2\Sigma^+$ with those of, say, CO, $X\,^1\Sigma^+$ on the basis of their total bond orders: ($1{\cdot}511 + 0{\cdot}500$) and ($1{\cdot}554 + 1{\cdot}000$) respectively. We shall see, however, that this is by no means the case.

THE STRUCTURE OF CO_2, $^1\Sigma_g^+$

This molecule and various states of its ion CO_2^+ have been discussed in detail by Mulliken (1935*b*), using qualitative molecular orbital considerations. Assuming the various K shell electrons to be localized on their respective atoms, and that the oxygen (2*s*) electrons are non-bonding, he described the remaining electrons on the basis of the configuration

$$CO_2,\ ^1\Sigma_g^+:\quad \ldots(\sigma_g)^2\,(\sigma_u)^2\,(\pi_u)^4\,(\pi_g)^4.$$

(See also Mulliken 1942, however.) The LCAO forms are

$$\psi(\sigma_g) = \delta^g_O\{\phi^g_O(2p\sigma) + \phi^g_{O'}(2p\sigma)\} + \delta^g_C\phi_C(2s),$$
$$\psi(\sigma_u) = \delta^u_O\{\phi^u_O(2p\sigma) - \phi^u_{O'}(2p\sigma)\} + \delta^u_C\phi_C(2p\sigma),$$
$$\psi(\pi_u) = \gamma^u_O\{\phi^u_O(2p\pi) + \phi^u_{O'}(2p\pi)\} + \gamma^u_C\phi_C(2p\pi),$$
$$\psi(\pi_g) = \gamma^g_O\{\phi^g_O(2p\pi) - \phi^g_{O'}(2p\pi)\}.$$

We shall assume the self-consistent fields to be such that we may, to good approximation, describe the σ-bonding by means of two localized molecular orbitals, each of which describes the σ-bonding between one of the oxygens and the carbon atom. That is, we assume that $\delta^g_O \approx \delta^u_O$, as is easily verified on writing out the corresponding Ψ^R_K's. The hybridization at the quadrivalent carbon atom will be diagonal; its valence state may be symbolized by C, V_4: $(K)\,(2t\sigma^*)\,(2t'\sigma^*)\,(2p\pi^*)^2$, where $(2t\sigma^*)$

points at O, and $(2t'\sigma^*)$ at O′. Using square brackets to denote localized molecular orbitals, we therefore write

$$\mathrm{CO_2},\ {}^1\Sigma_g^+:\quad \ldots [t\sigma]^2\,[t'\sigma]^2\,(\pi_u)^4\,[2p\pi_{\mathrm{O}}]^2\,[2p\pi_{\mathrm{O'}}]^2,$$

where $$\psi(t\sigma) = \gamma^t_{\mathrm{O}}\phi_{\mathrm{O}}(2p\sigma) + \gamma^t_{\mathrm{C}}\phi_{\mathrm{C}}(2t\sigma),$$

and similarly for $\psi(t'\sigma)$. In discussing overall symmetries, we retain Mulliken's configuration, but write

$$\psi(\sigma_g) = \frac{1}{\sqrt{2}}\{\psi(t\sigma)+\psi(t'\sigma)\},\quad \psi(\sigma_u) = \frac{1}{\sqrt{2}}\{\psi(t\sigma)-\psi(t'\sigma)\}.$$

The absolute electro-affinity of $\phi_{\mathrm{O}}(2p\sigma)$ is ~ 10 eV and that of

$$\phi_{\mathrm{C}}(2t\sigma) = \frac{1}{\sqrt{2}}\{\phi_{\mathrm{C}}(2s)+\phi_{\mathrm{C}}(2p\sigma)\}$$

may be estimated at $\sim 10{\cdot}4$ eV. Accordingly, we suppose that $\gamma^t_{\mathrm{O}} \approx \gamma^t_{\mathrm{C}}$. An analysis of the π electron system now shows that $\gamma^u_{\mathrm{O}} \approx \gamma^u_{\mathrm{C}}$: Any (π_u) electron moves in the field of two oxygens, each with charge $+\frac{1}{4}$, and one carbon atom, with charge $+\frac{1}{2}$ and, for $(2p\pi^*)$ orbitals, the term values associated with such atoms are very nearly equal. The electron distribution is therefore uniform, each atom having zero net charge. The π bond order is easily shown to be

$$p^{\pi}_{\mathrm{CO}} = \sqrt{2} \approx 1{\cdot}414$$

for each of the carbon-oxygen bonds. (If the symmetry of the molecule is disturbed, however, the charge distribution will be disturbed also; since $\mathrm{CO_2}$ absorbs strongly in the infra-red, this reorganization will be considerable.)

THE GROUND STATE $^2\Pi_g$ OF $\mathrm{CO_2^+}$

Consider, now, the ion $\mathrm{CO_2^+}$, $^2\Pi_g$ which is produced by the first ionization (13·7 eV, Price & Simpson 1939) of carbon dioxide. This ion is the lower state in the emission bands of $\mathrm{CO_2^+}$, which are observed at λ2900 and λ3660 (Buesno-Saullehi 1941; Mrozowski 1941, 1942). The ionization potential of a non-bonding $2p\pi$ oxygen electron is estimated at 14·7 eV (Mulliken 1934*a*), so that either (π_g) is antibonding, or the ion rearranges. Mulliken (1942), taking the former view, tentatively explained the increase in bond length which accompanies the ionization to the participation of the carbon atom's $3d\pi$ orbital in (π_g). This argument, though valid on grounds of symmetry, does not seem plausible energetically. We shall show that, on ionization of a (π_g) electron, the residual π electron distribution rearranges in such a way, that the π bond order decreases: $\mathrm{CO_2^+}$, $^2\Pi_g$ has the structure

$$\mathrm{CO_2^+},\ {}^2\Pi_g:\quad \ldots(\sigma_g)^2\,(\sigma_u)^2\,(\pi_u)^4\,(\pi_g)^3,$$

where we assume (σ_g) and (σ_u) to have the same forms as for the $\mathrm{CO_2}$ ground state. Setting $\gamma^u_{\mathrm{C}} = \sin\theta$, we find that

$$(3I_{\mathrm{O}}-2I_{\mathrm{C}}-2E_{\mathrm{C}}+E_{\mathrm{O}}) - 3\{2(I_{\mathrm{C}}-E_{\mathrm{C}})+(I_{\mathrm{O}}-E_{\mathrm{O}})\}\cos 2\theta + 8\sqrt{2}\,\beta_{\mathrm{CO}}\cot 2\theta = 0,$$

where $$I_{\mathrm{O}} = 17{\cdot}17,\quad E_{\mathrm{O}} = 2{\cdot}87,\quad I_{\mathrm{C}} = 11{\cdot}17,\quad E_{\mathrm{C}} = 0{\cdot}69\,\mathrm{eV},$$

and we take $\beta_{CO} = -3\,\text{eV}$ as usual. The π bond order becomes

$$p^{\pi}_{CO} = \sqrt{2}\sin 2\theta = 1{\cdot}378,$$

and the net charges are given by

$$Q_C = 2\cos 2\theta = +0{\cdot}436, \quad Q_O = \tfrac{1}{2} - \cos 2\theta = +0{\cdot}282 = Q_{O'}.$$

The reorganization of the ion is seen to be considerable (for without this we should have had $Q_C = 0$, $Q_O = +0{\cdot}500 = Q_{O'}$ and $p^{\pi}_{CO} = \sqrt{2}$).

THE FIRST EXCITED STATE $^2\Pi_u$ OF CO_2^+

The upper state of the CO_2^+ emission bands near $\lambda 3660$ has symmetry $^2\Pi_u$ and the structure

$$CO_2^+,\ ^2\Pi_u: \quad \ldots (\sigma_g)^2 (\sigma_u)^2 (\pi_u)^3 (\pi_g)^4;$$

it therefore arises from CO_2, $^1\Sigma_g^+$ by the removal of a bonding (π_u) electron. If $\gamma^u_C = \sin\theta$, as before, we find that

$$(I_O - 2I_C + E_O) - \{2(I_C - E_C) + (I_O - E_O)\}\cos 2\theta + 4\sqrt{2}\,\beta_{CO}\cot 2\theta = 0,$$

whence

$$p^{\pi}_{CO} = \tfrac{3}{4}\sqrt{2}\sin 2\theta = 1{\cdot}060,$$

$$Q_C = \tfrac{1}{2} + \tfrac{3}{2}\cos 2\theta = +0{\cdot}435, \quad Q_O = \tfrac{1}{4} - \tfrac{3}{4}\cos 2\theta = +0{\cdot}282 = Q_{O'}.$$

The π bond strength is considerably weaker in CO_2^+, $^2\Pi_u$, therefore, than in the $^2\Pi_g$ ground state.

THE SECOND EXCITED STATE $^2\Sigma_u^+$ OF CO_2^+

The identification of Henning's Rydberg series (1932) (of limit 18 eV), with the removal of a (σ_u) electron from CO_2, $^1\Sigma_g^+$ to leave $^2\Sigma_u^+$, the upper state of the $\lambda 2900$ emission bands of CO_2^+, is now well established. Let us discuss the structure of this excited ion, by attributing any reorganization to the more polarizable π electrons alone:

$$CO_2^+,\ ^2\Sigma_u^+: \quad \ldots (\sigma_g)^2 (\sigma_u) (\pi_u)^4 (\pi_g)^4.$$

By a now familiar argument the π bond order and charge distribution may be expressed in terms of the root $(0 \leqslant \theta \leqslant \tfrac{1}{2}\pi)$ of the equation

$$2(-2I_C + I_O + E_O) - 3\cos 2\theta\{2(I_C - E_C) + (I_O - E_O)\} + 8\sqrt{2}\,\beta_{CO}\cot 2\theta = 0,$$

$$p^{\pi}_{CO} = \sqrt{2}\sin 2\theta = 1{\cdot}413,$$

$$Q_C = \tfrac{1}{2} + 2\cos 2\theta = +0{\cdot}467, \quad Q_O = \tfrac{1}{4} - \cos 2\theta = +0{\cdot}266 = Q_{O'}.$$

The π bond order is therefore almost identical with that found for carbon dioxide itself, and the reorganization is negligible. In our localized molecular orbital approximation, however, each carbon-oxygen σ bond is only $\tfrac{3}{4}$ as strong as before the ionization.

BOND ORDERS AND BOND PROPERTIES

If our π bond order p^{π}_{CO} is to represent the strength of the π-bonding in a series of compounds, we should expect it to have the following properties: In such a series of CO bonds, p^{π}_{CO} should be a monotonically increasing function of the π bond energy.

It should therefore also represent the contribution of the π electrons to the first and second derivatives of the total zeroth-order energy with respect to the internuclear distances. So that, *ceteris paribus*, the bond distances should decrease and the force constants should increase monotonically with increasing p^{π}_{CO}. No such systematic correlation has been attempted for these molecules hitherto, since no methods were available for calculating the π bond orders.

Now, apart from CO^+, $B\,{}^2\Sigma^+$ and CO_2^+, ${}^2\Sigma_u^+$, all the systems which we have discussed above should furnish us with such a series of comparable CO bonds, if we disregard variations in the σ bonds (e.g. differences of hybridization at the various carbon atoms); in CO^+, $B\,{}^2\Sigma^+$ and CO_2^+, ${}^2\Sigma_u^+$ the σ-bonding is only $\frac{1}{2}$ and $\frac{3}{4}$ as strong, respectively. In order to avoid a detailed discussion of interaction constants in the triatomic molecules and speculations about dissociation energies (e.g. Herzberg 1942; Gaydon & Penney 1945), we shall compare the equilibrium internuclear separations in these systems in the light of the calculated π bond orders. To facilitate comparisons table 1 is constructed.

TABLE 1

molecule	state	p^{π}_{CO}	r_e (Å)	ref.
CO	$X\,{}^1\Sigma^+$	1·554	1·1284	H
	$A\,{}^1\Pi$	1·132	1·232	S
	$B\,{}^1\Sigma^+$	$\lesssim$ 1·838	1·118	S
CO^+	$X\,{}^2\Sigma^+$	1·838	1·115	H
	$A\,{}^2\Pi$	0·839	1·24	S
	$B\,{}^2\Sigma^+$	1·511 $(-\frac{1}{2}\sigma)$	1·168	S
CO_2	${}^1\Sigma_g^+$	1·414	1·163	M
CO_2	${}^2\Pi_g$	1·378	1·177	M
	${}^2\Pi_u$	1·060	1·228	M
	${}^2\Sigma_u^+$	1·413 $(-\frac{1}{4}\sigma)$	1·180	M

(The authorities referred to are H, Herzberg, 1939; S, Sponer, 1935; M, Mulliken, 1942.)

On plotting the r_e against the p^{π}_{CO}, the curve in figure 1 is obtained. In order of increasing p^{π}_{CO}, it is found that the six points

$$\mathrm{CO}^+,\ A\,{}^2\Pi;\quad \mathrm{CO}_2^+,\ {}^2\Pi_u;\quad \mathrm{CO}_2^+,\ {}^2\Pi_g;\quad \mathrm{CO}_2,\ {}^1\Sigma_g^+;\quad \mathrm{CO},\ X\,{}^1\Sigma^+;\quad \mathrm{CO}^+,\ X\,{}^2\Sigma^+$$

may all be connected by a smooth curve. Since the interaction of the Rydberg orbital ($r\sigma$) with the CO^+ core for CO, $B\,{}^1\Sigma^+$ was not amenable to direct calculation, we predicted $1{\cdot}554 < p^{\pi}_{\mathrm{CO}} \lesssim 1{\cdot}838$ for this excited state of carbon monoxide; from our curve it appears that we should take $p^{\pi}_{\mathrm{CO}} = 1{\cdot}76$, which is in complete harmony with our prediction. At low values of p^{π}_{CO} (near 1·0), we have deliberately avoided drawing the curve by some mean deviation method through CO^+, $A\,{}^2\Pi$; CO_2^+, ${}^2\Pi_u$ and CO, $A\,{}^1\Pi$, but have disregarded the last: This state contains five electrons in the sub-Rydberg π shell, which normally contains only four or less electrons, so that we should expect the exchange forces to give rise to an additional weakening which could not appear in our zeroth-order treatment. (Zeroth-order methods seek approximations to Hartree's equations; antisymmetrization is necessary before approximate solutions of the more accurate Fock equations can be obtained; for our purposes—we have not attempted energy calculations—the zeroth-order approximation should be sufficient, but will not disclose the exchange forces which tend to discourage the overcrowding of such sub-Rydberg shells.) In confirmation of this, CO, $A\,^1\Pi$ is found to lie above the curve.

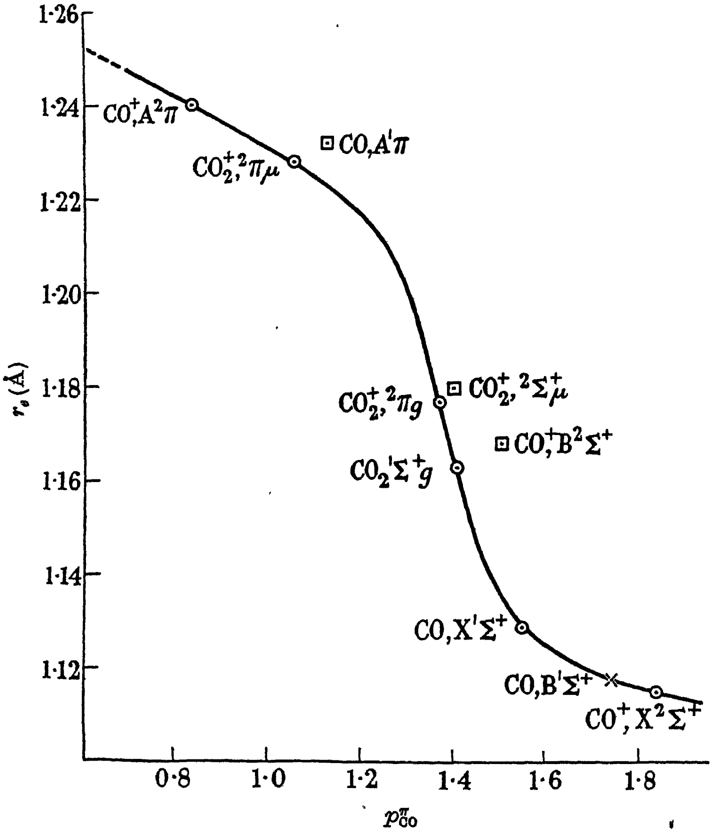


FIGURE 1. The internuclear distance/π bond order curve.

Our curve exhibits a possibly unexpected feature in its sharp descent and point of inflexion between 1·21 and 1·13 Å. Physically this implies that the π bond strength increases very rapidly in this range—a not unreasonable conclusion on simple overlap considerations. Since the carbon σ-bonding orbitals are 'diagonal', $\frac{1}{\sqrt{2}}(s+p)$, in CO_2^+, $^2\Pi_u$; most probably, predominantly p in CO^+, $A\,^2\Pi$; and near 'tetrahedral', $\frac{1}{2}(s+\sqrt{3}p)$, in CO, $X\,^1\Sigma^+$, it appears (as we have tacitly assumed) that variations in the type of σ-bonding are indeed relatively unimportant: That the first derivatives of the π electron energy terms are numerically considerably greater than those of the σ electrons—although the σ electron energy terms themselves will depend significantly on the hybridization ratios. In view of this, we should expect the bond lengthening, corresponding to quite large weakenings in the σ-bonding, to be relatively small at internuclear separations of less than about 1·25 Å. Our conclusion is confirmed by CO_2^+, $^2\Sigma_u^+$ and CO^+, $B\,^2\Sigma^+$: These are respectively only 0·017 and

0·034 Å longer than if they contained the usual two σ-bonding electrons instead of the $1\frac{1}{2}$ and 1 in these ions. (In our approximation, the as yet unknown state CO_2^+, ${}^2\Sigma_g^+$ arising from the removal of a (σ_g) electron from CO_2, ${}^1\Sigma_g^+$ should have the same moment of inertia as CO_2^+, ${}^2\Sigma_u^+$. In a higher approximation, which includes σ electron delocalization, it should be rather larger.)

Substantially the same results are obtained, if, instead of taking $|\beta_{CO}| = 3\,\text{eV}$, we were to use 2 or 4 eV for the numerical value of this integral, which is the only 'semi-empirical' parameter entering into our π electron analyses. And it seems fairly certain, *a priori*, that $|\beta_{CO}|$ must lie somewhere in this range. It may be objected that, since the π electron energy terms vary appreciably with changes in the internuclear distances, we should take into account the dependence of β_{CO} on the r_e's. But, on consideration, it is clear that in our series the resonance integrals are functions of the p^{π}_{CO} and therefore that no such refinements are required, providing we recognize that the π bond energy terms are not linear functions of the p^{π}_{CO}—and we have had no occasion to suppose this. It may, however, imply that our formal charges are slightly in error.

The author would like to thank Professor Coulson and Dr Barrow for helpful discussion and the Board of the British Rubber Producers' Research Association for a research grant.

REFERENCES

Buesno-Saullehi, F. 1941 *Phys. Rev.* **60**, 556.
Coulson, C. A. 1942 *Proc. Camb. Phil. Soc.* **38**, 210.
Coulson, C. A. 1948 *V. Henri Memorial Vol.* Liège: Desoer.
Coulson, C. A. & Moffitt, W. 1949 *Phil. Mag.* (in the press).
Dirac, P. A. M. 1935 *Quantum Mechanics.* Oxford Univ. Press.
Gaydon, A. G. & Penney, W. G. 1945 *Proc. Roy. Soc.* A, **183**, 374.
Henning, H. J. 1932 *Ann. Phys., Lpz.*, (5), **13**, 599.
Herzberg, G. 1939 *Molecular structure.* New York: Prentice Hall.
Herzberg, G. 1942 *J. Chem. Phys.* **10**, 306.
Hund, F. 1933 *Handb. Phys.* **24**, 561.
Long, L. H. & Walsh, A. D. 1947 *Trans. Faraday Soc.* **43**, 342.
Matheson, L. A. 1932 *Phys. Rev.* **40**, 813.
Moffitt, W. 1948 *Proc. Phys. Soc.* **60**, 597.
Moffitt, W. 1949 *Proc. Roy. Soc.* A, **196**, 510.
Moffitt, W. & Coulson, C. A. 1947 *Phil. Mag.* **38**, 634.
Mrozowski, S. 1941 *Phys. Rev.* **60**, 730.
Mrozowski, S. 1942 *Phys. Rev.* **62**, 720.
Mulliken, R. S. 1928 *Phys. Rev.* **32**, 205.
Mulliken, R. S. 1932 *Rev. Mod. Phys.* **4**, 1.
Mulliken, R. S. 1934*a* *J. Chem. Phys.* **2**, 782.
Mulliken, R. S. 1934*b* *J. Chem. Phys.* **2**, 400.
Mulliken, R. S. 1935*a* *J. Chem. Phys.* **3**, 573.
Mulliken, R. S. 1935*b* *J. Chem. Phys.* **3**, 720.
Mulliken, R. S. 1942 *Rev. Mod. Phys.* **14**, 204.
Pauling, L. 1932 *J. Amer. Chem. Soc.* **54**, 998.
Price, W. C. & Simpson, D. M. 1939 *Proc. Roy. Soc.* A, **169**, 501.
Slater, J. C. 1930 *Phys. Rev.* **36**, 56.
Sponer, H. 1935 *Molekülspektren.* Berlin: Springer.
Van Vleck, J. H. 1934 *J. Chem. Phys.* **2**, 20, 297.
Zener, C. 1930 *Phys. Rev.* **36**, 51.

The kinetics of certain reactions between methyl halides and anions in water

By E. A. Moelwyn-Hughes

Department of Physical Chemistry, University of Cambridge

(*Communicated by R. G. W. Norrish, F.R.S.—Received* 12 *August* 1948—*Revised* 4 *November* 1948)

The rates at which methyl chloride, methyl bromide and methyl iodide react with the hydroxyl and the thiosulphate ions in water have been measured at various concentrations and temperatures. The apparent energies of activation in both series increase in the same direction as the dipole moment of the methyl halide. The results are discussed in terms of a theory of the kinetics of the reactions between ions and polar molecules in solution.

Introduction

It is now generally agreed that a classical electrostatic energy, E_e, forms part of the total energy which must be acquired by ionic or polar reactants before they can undergo chemical change in solution. On this view, the apparent energy of activation, E_A, for reactions between ions and polar molecules should, in general, be proportional to the electrical moment, μ_A, of the polar group concerned. Direct experimental evidence on this point seems, however, to be lacking, though relationships exist (Nathan & Watson 1933) between E_A and the polarity of groups substituted in the molecule at points remote from the seat of ionic attack. The most that can be said from a review of the literature is that the molar energies of activation for reactions of alkyl chlorides in hydroxylic media exceed, by about 1 kcal., the energies of activation for reactions of the corresponding alkyl iodides with a common ion. Such is the case, for example, in ethyl-alcoholic solution, with allyl halides reacting with the ethoxide ion (Conrad & Brückner 1889) and with ethyl halides reacting with the hydroxyl ion (Grant & Hinshelwood 1933). Within the limits of accuracy claimed in these experiments, no difference can be detected between E_A for the bromide and the iodide. It is true that some early work on the kinetics of the reaction of methyl halides with the thiosulphate ion in water (Slator 1904) can be shown to reveal a gradual decrease in E_A as we ascend the series, chloride, bromide and iodide, which is also the order of decreasing polarity, but the velocity constants upon which the values of E_A rest, have, except in the case of methyl chloride, been measured at two temperatures only, differing by 10°. Moreover, Slator's data on the reactions of the same ion with the three ethyl-halogeno-acetates fail to reveal any difference between the energies of activation of the bromide and iodide reactions.

The object of the experiments described here is to measure as precisely as possible the energies of activation of the reactions of the methyl halides with the hydroxyl ion,

$$CH_3X + OH^- \rightarrow CH_3OH + X^-,$$

in aqueous solution, and to inquire into the way in which they may be related to the dipole moments. Slator's reactions,

$$CH_3X + S_2O_3^{--} \rightarrow CH_3S_2O_3^{-} + X^{-},$$

have also been repeated over wider experimental conditions, yielding results which confirm his.

EXPERIMENTAL METHOD

Solutions of methyl chloride, methyl bromide and methyl iodide were prepared by passing the vapours through water. To these, after heating, were added aqueous solutions of sodium hydroxide or thiosulphate, and the reaction mixture was quickly drawn into a thermostatic, vapour-free reactor which has been previously described (Moelwyn-Hughes 1941). Known volumes of the solution were expressed at intervals, chilled and analyzed.

In the first set of reactions, the rates were measured by estimating the alkali consumed and the halogen ion produced. Bromocresol blue was used as indicator in estimating the base, and Volhard's method was used for estimating inorganic halide. The reaction was followed for a time ten times as great as the time of half-completion. Generally, the initial concentration a of methyl halide employed was between 30 and 30 millimoles per litre, and the initial concentration b of alkali was, in most experiments, from two to three times as great.

In the second set of reactions, the rate with methyl iodide was measured as previously described for methyl bromide (Moelwyn-Hughes 1941), but most of the runs with methyl chloride were done by removing samples of the reaction solution from beneath undisturbed films of dodecyl alcohol.

KINETIC ANALYSIS OF EXPERIMENTS WITH THE HYDROXYL ION

Methyl halides in aqueous alkaline solution react not only with the hydroxyl ion but with the water molecules also. The latter reaction is known to be unimolecular, and the former reaction may be assumed on good grounds to be bimolecular. If x denotes the concentration of organic halide which has reacted in time t, we have the following kinetic scheme:

$$\underset{(a-x)}{CH_3X} + H_2O \xrightarrow{k_1} CH_3OH + H^{+} + X^{-}, \tag{1}$$

$$\underset{(a-x)}{CH_3X} + \underset{(b-x)}{OH^{-}} \xrightarrow{k_2} CH_3OH + X^{-}. \tag{2}$$

The instantaneous rate of reaction is then given by the differential equation

$$\frac{dx}{dt} = k_1(a-x) + k_2(a-x)(b-x), \tag{3}$$

which gives, on integration,

$$t = \frac{1}{k_1 + k_2(b-a)} \ln\left[\left(\frac{a}{a-x}\right)\frac{k_1 + k_2(b-x)}{k_1 + k_2 b}\right]. \tag{4}$$

This expression naturally reduces to the ordinary laws of unimolecular and bimolecular processes when k_2 and k_1 are, respectively, zero. The fractional change at time t is seen to be

$$\frac{x}{a} = \frac{e^{\alpha t}-1}{e^{\alpha t}-\beta}, \tag{5}$$

where

$$\alpha = k_1 + k_2(b-a) \tag{6}$$

and

$$\beta = k_2 a/(k_1 + k_2 b). \tag{7}$$

Conditions of temperature and concentration must be rather carefully chosen in order to allow an accurate evaluation of both k_1 and k_2 from the chemical analyses of a single run. The method is described in a later section. For the present, we note that values of k_1 are already known for the three methyl halides at all accessible temperatures (Moelwyn-Hughes 1938*a*). It is therefore convenient to define a drifting term k_2' by means of the relation

$$\frac{dx}{dt} = k_2'(a-x)\,(b \quad x). \tag{8}$$

TABLE 1. BIMOLECULAR VELOCITY CONSTANTS FOR THE REACTIONS OF THE METHYL HALIDES WITH THE HYDROXYL ION IN WATER

t (min.)	millimoles/l. $(a-x)$	millimoles/l. $(b-x)$	$k_2 \times 10^5$ (l./mole-sec.)	
	CH_3Cl at 310·94° K.	$k_1 = 1{\cdot}27 \times 10^{-7}$ sec.$^{-1}$		
0	33·3	99·5	—	
1000	27·0	93·2	3·49	3·55 ± 0·06
2000	22·0	88·2	3·56	
3000	18·2	84·4	3·54	
4000	15·1	81·3	3·55	
5000	12·5	78·7	3·59	
∞	0	66·2	—	
	CH_3Br at 308·46° K.	$k_1 = 1{\cdot}465 \times 10^{-6}$ sec.$^{-1}$		
0	62·8	104·9	—	
50	53·3	95·4	53·3	53·3 ± 1·0
100	45·8	87·9	53·4	
175	36·9	79·0	54·3	
250	30·9	73·0	52·9	
350	24·4	66·5	53·2	
450	19·9	62·0	52·5	
∞	0	42·1	—	
	CH_3I at 342·80° K.	$k_1 = 2{\cdot}45 \times 10^{-5}$ sec.$^{-1}$		
0	11·76	103·60	—	
3·5	9·79	101·63	(828)	855 ± 12
8·5	7·52	99·36	841	
13·5	5·57	97·41	(897)	
19·5	4·20	96·04	863	
26·5	2·92	94·76	867	
37·5	1·75	93·59	847	
∞	0	91·84	—	

Comparison with equation (3) shows that the true bimolecular constant is thus

$$k_2 = k_2' - k_1/(b-x). \tag{9}$$

Representative values found in this way are given in table 1. Under favourable circumstances, the constants found in a given run did not vary by more than 1 % from the mean value. For most runs, the deviations were about ± 2 %. The two bracketed values indicate titration errors of 0·05 and 0·19 ml., respectively, of standard (0·02 N) reagent per 20 ml. sample of solution. The degree of accuracy of the results, however, is better judged by calculating the percentage change according to equation (5). Typical results are given in table 2. Had no correction been made for the interaction with the solvent, higher values of k_2 would have been found.

TABLE 2. FRACTIONAL CHANGES OF THE METHYL HALIDES IN AQUEOUS ALKALINE SOLUTION(k_1 and α in sec.$^{-1}$)

t	millimoles/l.		100(x/a)	
(min.)	($a-x$)	($b-x$)	observed	calculated
CH_3Cl at 333·05° K: $k_1 = 2{\cdot}16 \times 10^{-6}$, $k_2 = 4{\cdot}88 \times 10^{-4}$, $\alpha = 3{\cdot}50 \times 10^{-5}$, $\beta = 3{\cdot}30 \times 10^{-1}$				
0	35·3	102·5	0	0
2	35·1	102·3	0·6	0·6
30	32·0	99·2	9·3	8·8
65	29·4	96·6	16·7	18·0
109	25·5	92·7	27·7	27·7
214	18·9	86·1	46·5	45·8
1160	2·1	69·3	94·0	93·9
∞	0	67·2	100·0	100·0
CH_3Br at 297·62° K: $k_1 = 3{\cdot}21 \times 10^{-7}$, $k_2 = 1{\cdot}373 \times 10^{-4}$, $\alpha = 6{\cdot}88 \times 10^{-6}$, $\beta = 5{\cdot}33 \times 10^{-1}$				
0	57·2	105·0	0	0
200	48·4	96·2	15·4	15·5
500	38·5	86·3	32·7	32·9
900	29·0	76·8	49·3	49·1
1400	21·1	68·9	63·1	62·7
2000	15·2	63·0	73·4	72·5
3000	8·9	56·7	84·4	84·4
∞	0	47·8	100·0	100·0
CH_3I at 342·80° K: $k_1 = 2{\cdot}45 \times 10^{-5}$, $k_2 = 8{\cdot}55 \times 10^{-3}$, $\alpha = 8{\cdot}10 \times 10^{-4}$, $\beta = 1{\cdot}105 \times 10^{-1}$				
0	11·76	103·60	0	0
3·5	9·79	101·63	16·8	17·2
8·5	7·52	99·36	36·0	36·5
13·5	5·57	97·41	50·7	51·0
19·5	4·20	96·04	64·3	64·0
26·5	2·92	94·76	75·2	74·7
37·5	1·75	93·59	85·1	85·3
∞	0	91·84	100·0	100·0

EXPERIMENTS WITH EXCESS OF METHYL HALIDE

It has been assumed in this treatment that the hydrogen ion formed by reaction (1) instantly reacts with the hydroxyl ion. Equations (4) and (5) consequently lose their physical significance when a is greater than b. An evaluation of the constants

k_1 and k_2 from analyses of systems containing higher concentrations of organic halide than of alkali is, however, not excluded. In such experiments, the early part of the reaction, while the solution is still alkaline, is rapid. After the neutrality point, it becomes slow, and indistinguishable from the hydrolysis of the methyl halide in water containing some salt. The results of the early portion of a typical experiment, in which 10 ml. samples were analyzed for alkali and halide ion, are given in table 3.

TABLE 3. THE REACTION OF SODIUM HYDROXIDE WITH EXCESS OF METHYL IODIDE AT 342·81° K

t (min.)	uncorrected thermometer reading (° C)	(millimoles/l.) $(a-x)$	$(b-x)$	$(a-b)$	$k_1' \times 10^5$ (sec.$^{-1}$)
0	69·86	46·40	19·24	27·16	—
13	69·80	40·61	13·44	27·17	14·16
25·5	69·84	36·83	9·60	27·23	10·81
44	69·86	33·02	6·04	26·98	8·02
67	69·90	30·20	3·44	26·76	5·67
100	69·84	28·38	1·04	27·34	3·56
∞	—	0	0	—	2·64

The most convenient method of evaluating the constants is to define a drifting term, k_1', by the relation

$$\frac{dx}{dt} = k_1'(a-x). \tag{10}$$

Comparison with equation (3) shows that

$$k_1' = k_1 + k_2(b-x). \tag{11}$$

The values given in the last column of table 3 conform with this linear relationship. The intercept gives $k_1 = 2\cdot64 \times 10^{-5}$ sec.$^{-1}$, which compares favourably with that determined directly and in the absence of alkali, which at this temperature is $2\cdot56 \times 10^{-5}$ sec.$^{-1}$ (Moelwyn-Hughes 1938*a*). From the gradient of equation (11), we obtain $k_2 = 8\cdot56 \times 10^{-3}$ l./mole-sec. It is worth noting that, on account of a zero time error, such concordance in the results could not have been achieved had we evaluated the constants as a function of time rather than of concentration.

Experiments in this category were carried out in order to see whether the hydrolytic reaction retains its unimolecular character in the presence of alkali, and to measure the effect of concentration of organic halide on the rate. The first point seems to be well enough established. By varying the concentration of methyl iodide as far as is possible between the upper limit imposed by its solubility and the lower limit consistent with reliable analysis, no effect due to the concentration of methyl halide could be detected (table 4) by the present technique. In the region covered, a four-fold change in the concentration of either reactant does not affect the velocity coefficient by more than 4 %. This finding is in agreement with anticipation, for great dilution effects are uncommon in the case of reactions between ions and polar molecules in hydroxylic media. The small effect is now being investigated by a conductivity method.

TABLE 4. THE INFLUENCE OF DILUTION ON THE KINETICS OF THE REACTION BETWEEN METHYL IODIDE AND THE HYDROXYL ION IN WATER

	millimoles/l.		
T (° K)	$a = [CH_3I]_0$	$b = [OH^-]_0$	$k_2 \times 10^3$ (l./mole-sec.)
342·85	39·5	103·6	8·53 ± 0·39
342·85	25·1	103·6	8·51 ± 0·18
342·80	11·8	103·6	8·55 ± 0·12
342·81	46·4	19·2	8·56 ± 0·21

DETERMINATION OF THE ENERGIES OF ACTIVATION

The apparent energy of activation, defined by the Arrhenius equation

$$E_A = RT^2\left(\frac{d\ln k}{dT}\right)_P = -R\left[\frac{d\ln k}{d(1/T)}\right]_P, \qquad (12)$$

is found to be a constant for each of the reactions studied, within the accuracy and temperature range of the experiments. The data can thus be represented in the form

$$k_2 = A\,\mathrm{e}^{-E_A/RT}, \qquad (13)$$

where A is a second constant. The applicability of this relation is shown in table 5. The numerical values of A and E_A are given in a later section (table 9).

TABLE 5. BIMOLECULAR VELOCITY CONSTANTS FOR THE REACTIONS OF THE METHYL HALIDES WITH THE HYDROXYL ION IN WATER AT VARIOUS TEMPERATURES

	$k_2 \times 10^4$ (l./mole-sec.)	
T (° K)	observed	calculated (equation (13))
	$CH_3Cl + OH^- \rightarrow CH_3OH + Cl^-$	
310·94	0·355	0·358
322·46	1·46	1·46
328·61	3·00	2·98
333·05	4·88	4·83
339·68	9·97	9·86
	$CH_3Br + OH^- \rightarrow CH_3OH + Br^-$	
292·69	0·708	0·703
297·62	1·37	1·37
303·41	2·87	2·86
308·46	5·33	5·37
313·29	9·31	9·48
322·91	29·1	28·6
	$CH_3I + OH^- \rightarrow CH_3OH + I^-$	
303·00	1·16	1·16
312·23	3·45	3·49
314·79	4·52	4·61
322·03	10·6	10·3
333·03	32·1	32·6
342·83	85·4	83·8

REACTIONS OF THE METHYL HALIDES WITH THE THIOSULPHATE ION

With the concentrations of halide and salt used in this work, which were roughly 20 and 50 millimoles/l., respectively, the extent of reaction with the solvent is slight. The exact value at 317·24° K, for example, is 1·2 % of the halide initially present. For this reason, the side reaction can be ignored, and equations (6), (7) and (5) assume simpler forms:

$$\alpha = k_2(b-a), \tag{14}$$

$$\beta = a/b, \tag{15}$$

and

$$\frac{x}{a} = \frac{e^{k_2(b-a)t}-1}{e^{k_2(b-a)}-(a/b)}. \tag{16}$$

TABLE 6. THE KINETICS OF THE REACTION BETWEEN METHYL IODIDE AND SODIUM THIOSULPHATE

$T = 292{\cdot}28^\circ$ K: $k_2 = 1{\cdot}56 \times 10^{-2}$ l./mole-sec.; $\alpha = 3{\cdot}70 \times 10^{-4}$ sec.$^{-1}$, $\beta = 0{\cdot}579$

t	millimoles/l.		$k_2 \times 10^2$	$100(x/a)$	
(min.)	$(a-x)$	$(b-x)$	(l./mole-sec.)	observed	calculated
0	32·6	56·3	—	0	0
4	26·7	50·4	1·56	18·1	17·9
9	21·43	45·13	1·55	34·3	34·6
13	18·45	42·15	1·51	43·4	44·2
22	12·97	36·67	1·58	60·2	59·9
31·5	9·60	33·30	1·57	70·6	70·6
44	6·34	30·04	1·61	80·6	79·7
64	3·82	27·52	1·57	88·3	88·1
∞	0	23·70	—	100·0	100·0

TABLE 7. BIMOLECULAR VELOCITY CONSTANTS FOR THE REACTIONS OF THE METHYL HALIDES WITH THE THIOSULPHATE ION AT VARIOUS TEMPERATURES

	$k_2 \times 10^4$ (l./mole-sec.)	
T (° K)	observed	calculated (equation (13))
	$CH_3Cl + S_2O_3^= \rightarrow CH_3S_2O_3^- + Cl^-$	
293·14	4·96	5·12
298·04	9·09	9·06
303·07	16·2	16·3
313·03	47·1	47·6
322·70	132	130
	$CH_3Br + S_2O_3^= \rightarrow CH_3S_2O_3^- + Br^-$	
273·23	19·3	19·2
277·94	33·6	34·2
282·26	60·0	59·4
288·56	130	130
293·10	219	216
298·17	377	375
	$CH_3I + S_2O_3^= \rightarrow CH_3S_2O_3^- + I^-$	
282·26	48·3	48·2
286·02	75·6	76·4
292·28	156	156
298·59	315	311

As the bimolecularity of this reaction has already been proved (Slator 1904), only one set of data is given, to illustrate the kind of result obtained by the different technique adopted in the present work (table 6). The more extensive data on the temperature effect, however, are summarized in full, as this was the object of the repetition (table 7). The two sets of results agree as well with each other as the difference in the experimental procedures allows.

THE DIRECT DETERMINATION OF SMALL-ENERGY DIFFERENCES

Because the difference between the energies of activation is relatively small, the following scheme has been tried, in order to eliminate the errors arising from the measurement of time and temperature, leaving only errors due to chemical analysis. A solution containing roughly the same concentrations of both methyl bromide and methyl iodide was allowed to react with more than sufficient alkali to convert both into the alcohol. If, after a time t, x moles of bromide ion have been formed from an initial concentration of a moles of methyl bromide, and after the same time, y moles of iodide ion have been formed from an initial concentration of b moles of methyl iodide, the instantaneous rates of reaction are

$$\frac{dx}{dt} = k_2(a-x)(c-x-y) \quad \text{and} \quad \frac{dy}{dt} = k_4(b-y)(c-x-y),$$

where c is the initial concentration of hydroxyl ion. The experiments must be performed under conditions where the reactions with the solvent are unimportant. The following functional relationship between the concentrations should then hold at all times:

$$\frac{k_2}{k_4} = \frac{\ln[a/(a-x)]}{\ln[b/(b-y)]}, \tag{17}$$

from which

$$\Delta E_A = E_2 - E_4 = -R\left[\frac{d\ln(k_2/k_4)}{d(1/T)}\right]_P. \tag{18}$$

On analyzing reaction samples for iodide ion and total inorganic halide and testing equation (17), it was found that the ratio k_2/k_4 increased steadily during the run until it finally becomes almost double its initial value. The values of the ratio extrapolated to zero concentrations of x and y compare well with those obtained from runs where the halides were allowed to react separately with alkali (see table 8), and indicate a value of about 400 cal. for ΔE_A, which is the lower limit according to the summarized results given in table 9. The failure of equation (17) in the present case is due to the interaction of the products, according to the scheme

$$CH_3I + Br^- \rightleftharpoons CH_3Br + I^-,$$

of which the forward reaction is faster than, and the reverse reactions lower than those under investigation (Moelwyn-Hughes 1938*b*). Owing to this circumstance, the method is deprived of its accuracy in the present case, but it is proving helpful in other experiments where the attacking ion replaces both halogen atoms more rapidly than they can replace each other.

TABLE 8. THE COMPARATIVE RATES OF REACTION OF METHYL BROMIDE AND METHYL IODIDE WITH THE HYDROXYL ION

	k_2/k_4	
T (°K)	from separate runs	from runs with mixed halides
302·51	2·29	2·32
312·70	2·39	2·37

TABLE 9. SUMMARY OF EXPERIMENTAL RESULTS, WITH THOSE OF SLATOR SHOWN IN BRACKETS

	l./mole-sec.		E_A
reaction	$k_{298\cdot1^\circ \mathrm{K}}$	$A \times 10^{-12}$	(cal./mole)
$CH_3Cl + OH^- \rightarrow CH_3OH + Cl^-$	$6\cdot67 \times 10^{-6}$	4·11	$24{,}280 \pm 120$
$CH_3Br + OH^- \rightarrow CH_3OH + Br^-$	$1\cdot435 \times 10^{-4}$	10·4	$23{,}000 \pm 150$
$CH_3I + OH^- \rightarrow CH_3OH + I^-$	$6\cdot36 \times 10^{-5}$	1·24	$22{,}220 \pm 230$
$CH_3Cl + S_2O_3^{--} \rightarrow CH_3S_2O_3^- + Cl^-$	$9\cdot14 \times 10^{-4}$	1·01	20,520
	$(8\cdot37 \times 10^{-4})$	(1·03)	(20,580)
$CH_3Br + S_2O_3^{--} \rightarrow CH_3S_2O_3^- + Br^-$	$3\cdot24 \times 10^{-2}$	6·82	19,540
	$(3\cdot47 \times 10^{-2})$	(4·68)	(19,280)
$CH_3I + S_2O_3^{--} \rightarrow CH_3S_2O_3^- + I^-$	$2\cdot84 \times 10^{-2}$	2·01	18,880
	$(3\cdot31 \times 10^{-2})$	(2·29)	(18,880)

DISCUSSION

These experiments establish the fact that, in the substitution of the hydroxyl ion for the halogen atom X in the reactions

$$OH^- + CH_3X \rightarrow CH_3OH + X^-,$$

the energy of activation in aqueous solution increases as the dipole moment of the methyl halide increases (figure 1). Moreover, as could have been anticipated, the increase is greater for the smaller molecules. Ignoring this difference, and allowing the widest scope to the errors of experiment, the results are broadly summarized by the equation

$$\frac{dE_A\,(\text{kcal.})}{d\mu_A\,(\text{Debye units})} = 10\cdot7 \pm 8\cdot6. \tag{19}$$

In terms of a theory of electrokinetic effects attending reactions between ions and polar molecules in solution (Moelwyn-Hughes 1936), the bimolecular velocity constant is given by the formula:

$$k_2 = Z\,e^{-E_n/RT}\,e^{-E_e/RT}, \tag{20}$$

where Z is the number of collisions between ions and molecules per second in unit volume of solution when they are present at a standard concentration. E_n and E_e are, respectively, the non-electrostatic and the electrostatic components of the energy of activation. As the latter varies inversely with respect to the dielectric constant D which generally depends on temperature according to Abegg's law,

$$-\left(\frac{d\ln D}{dT}\right)_P = L, \tag{21}$$

it follows from equation (12) that the observed energy of activation is

$$E_A = E_n + (1 - LT)\, E_e. \tag{22}$$

The derivation implies that Z and E_n are independent of temperature. E_e is given in terms of the valency z_B of the ion, the dipole moment μ_A of the reacting molecule, their distance apart a_c in the critical configuration, and the angle θ subtended by the ion from the positive end of the dipole:

$$E_e = N_0 z_B \epsilon \mu_A \cos\theta / D a_c^2. \tag{23}$$

N_0 is the Avogadro number, and ϵ the electronic charge. Hence,

$$E_A = E_n + (1 - LT)\, N_0 z_B \epsilon \mu_A \cos\theta / D a_c^2. \tag{24}$$

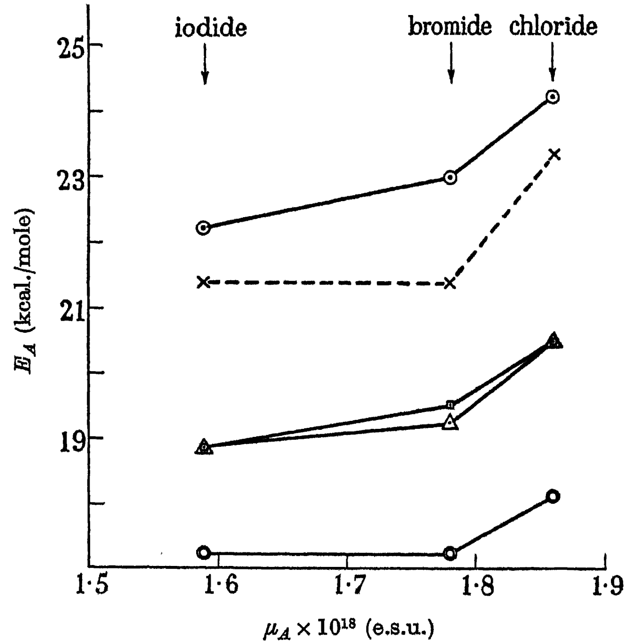


FIGURE 1. Apparent energies of activation for reactions of organic halides as a function of their dipole moments.

⊙ Methyl halides with OH^- in water (this work).
× Ethyl halides with OH^- in ethanol (Grant & Hinshelwood).
⊡ Methyl halides with $S_2O_3^{--}$ in water (this work).
△ Methyl halides with $S_2O_3^{--}$ in water (Slator).
◎ Allyl halides with $C_2H_5O^-$ in ethanol (Conrad & Brückner).

General arguments based on the optical inversion which accompanies such substitutions at an asymmetrically saturated carbon atom (Bergmann, Polanyi & Szabo 1933) suggest that θ in the present reactions is less than 90°, i.e. that the ion approaches the positive end of the molecule. Cos θ may thus be assumed positive.

If we attribute the change in E_A to electrostatic effects, and thus regard E_n as constant, there are two courses open, according to the choice of dielectric constant. If, in the first place, we accept the value for water, the term $(1 - LT)$ at 25° C is found to be $-0{\cdot}38$. Since z_B is -1, the second term in equation (24) becomes positive, in agreement with experiment. Its magnitude, however, is so low (0·08 kcal.) as to render it insignificant. The numerical estimate is based on a value of 2·75 Å for a_c, which, determined kinetically in previous work on the reaction between methyl bromide and the thiosulphate ion, is probably too high. If, in the second place, we take D as unity, and L consequently as zero, the second term of equation (24), using the same value of a_c, becomes $-16{\cdot}28$ kcal., which is of the right order of magnitude but of the wrong sign. We are thus faced with the same dilemma as has appeared in a study of the kinetics of reactions between polar molecules in non-ionizing media, where the electrostatic contribution to the energy of activation is positive though, on chemical grounds, we have every reason to suppose that the reactants aline themselves in such positions that the interaction energy is negative (Moelwyn-Hughes 1947). The method of resolution attempted in that case is, it is thought, sufficiently plausible to justify its application to the present problem.

Let it be supposed that, contrary to what is admittedly required in a complete formulation, the total energy u of interaction between the ion and the dipole can be represented as a function of their separation a according to the equation†

$$u = Aa^{-9} - Ba^{-2}. \tag{25}$$

The equilibrium separation a_e and the energy u_e of the system corresponding to it, are obtained from the condition governing a minimum energy, and are given by the relations

$$a_e^7 = 9A/2B, \tag{26}$$

and

$$u_e = -7B/9a_e^2. \tag{27}$$

By eliminating A and B, equation (25) may be written in the form

$$\frac{u}{u_e} = \tfrac{1}{7}[9(a_e/a)^2 - 2(a_e/a)^9]. \tag{28}$$

A plot of $1-(u/u_e)$ as a function of $1-(a_e/a)$ is shown in figure 2.

A datum from extra-kinetic sources is required for the numerical evaluations, and may be obtained by accepting the thermal value for the minimum energy of interaction u_e of the hydroxyl ion with a single water molecule. Because of the difficulty in computing the lattice energy of the alkalies, this value is not known directly. The fluoride ion, however, is structurally so similar to the hydroxyl ion that we can assume the heats of hydration of both to have a common value. If the co-ordination number is 6, then the value required (u_e) is $-(\frac{1}{6})$ times the molar heat evolved during the dissolution of the fluoride ion in water. We thus have $-N_0u_e = 20{\cdot}7$ kcal./gram-ion (Latimer, Pitzer & Slansky 1939). On combining this datum with the theoretical expression for the attraction constant, which, when θ is zero, is

$$B = \epsilon\mu_A, \tag{29}$$

the repulsion constant is found to be

$$A = 4{\cdot}51 \times 10^{-82}\,\text{erg-cm.}^9, \tag{30}$$

---

† The symbol A in this equation, and subsequently, denotes the intrinsic repulsion constant, and is not to be confused with the Arrhenius constant of equation (13).

which compares well with the values of $3{\cdot}50 \times 10^{-82}$ and $4{\cdot}00 \times 10^{-82}$ obtained, respectively, for the Ne ... Ne pair from data on gases and for the F^- ... Na^+ from data on crystals (Lennard-Jones 1936). As the repulsion in all cases is exerted between the hydroxyl ion and the methyl group of atoms, this value of A is adopted for the three methyl halides. Equilibrium distances a_e and minimal energies u_e are now readily obtained (table 10). Similarly, by identifying the interaction energy u_c of the critical system with the experimental figure for E_A/N_0, the critical separations a_c are found (see arrow in figure 2). They are seen to be about $\frac{3}{4}$ times the equilibrium separations.

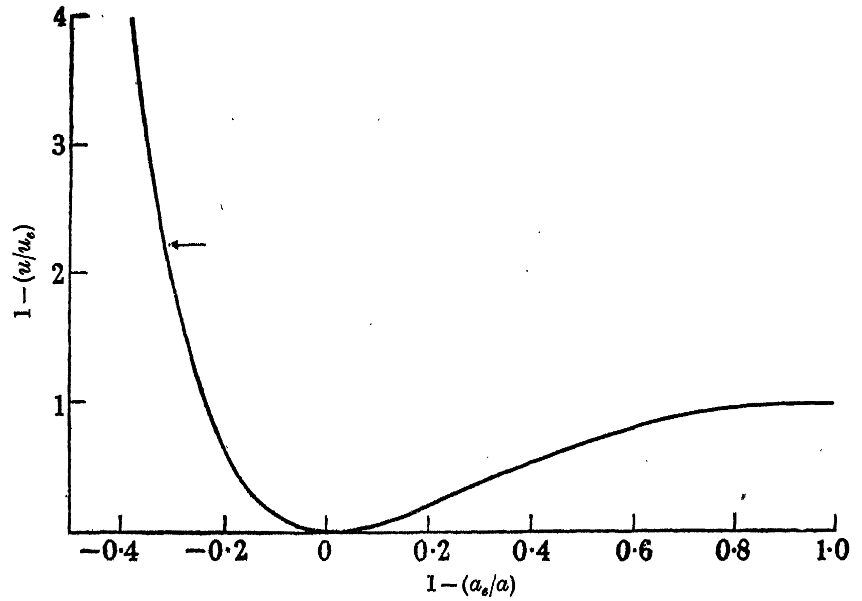


FIGURE 2. $u/u_e = \frac{1}{7}[9(a_e/a)^2 - 2(a_e/a)^9]$.

TABLE 10. SOME CONSTANTS OF ION-DIPOLE SYSTEMS IN STATES OF MINIMAL AND OF CRITICAL ENERGIES

	CH_3Cl	CH_3Br	CH_3I
$\mu_A \times 10^{18}$ (e.s.u.)	1·86	1·78	1·59
u_e (kcal.)	−20·70	−19·57	−16·91
$a_e \times 10^8$ (cm.)	2·17	2·19	2·22
u_c (kcal.)	24·28	23·00	22·22
$a_c \times 10^8$ (cm.)	1·65	1·66	1·68
u_c/u_e	−1·172	−1·175	−1·314
a_e/a_c	1·316	1·316	1·323
$f(a_e/a_c)$	0·896	0·896	1·011
$Z \times 10^{12}$ (l./mole-sec.)	1·59	1·43	1·25

By combining equations (27), (28) and (29), the energy of the critical complex may be expressed in the form

$$u_c = \frac{\epsilon\mu_A}{a_e^2}\left[\frac{2}{9}\left(\frac{a_e}{a_c}\right)^9 - \left(\frac{a_e}{a_c}\right)^2\right]. \tag{31}$$

The term in the square brackets, which is given in table 10 as $f(a_e/a_c)$, lies in the region 0·95 ± 0·05 for the three reactions; hence,

$$u_c = 0{\cdot}95\epsilon\mu_A/a_e^2. \tag{32}$$

On this basis, we obtain, after a conversion of units,

$$\frac{dE_A\,(\text{kcal.})}{d\mu_A\,(\text{Debye units})} = 13{\cdot}6 \pm 0{\cdot}6, \tag{33}$$

which, in respect of magnitude and sign, agrees with experiment. It may be incidentally noted that $u_c/|-u_e|$ is approximately 1·22 for each system.

The number of collisions, under standard conditions, between molecules and ions of reduced mass m^* approaching each other with relative velocities exceeding the critical value, may be taken as

$$Z = a_c^2(8\pi kT/m^*)^{\frac{1}{2}}\left(1+\frac{u_c}{kT}\right) \doteqdot a_c^2 u_c(8\pi/m^*kT)^{\frac{1}{2}}. \tag{34}$$

The special form assumed by utilizing equation (32) is

$$Z = 0{\cdot}95\epsilon\mu_A\left(\frac{a_c}{a_e}\right)^2\left(\frac{8\pi}{m^*kT}\right)^{\frac{1}{2}} = \frac{2{\cdot}74\epsilon\mu_A}{(m^*kT)^{\frac{1}{2}}}. \tag{35}$$

Numerical values of Z calculated by means of the last equation are given in table 10. They are in reasonable agreement with the experimental values of table 9.

Reference may be made here to the effect of the same polar groups on the energies of activation of reactions between the hydroxyl ion and *para*-substituted ethyl benzoates in aqueous alcoholic solution, which, compared with the present values of $dE_A/d\mu_A$, is smaller by a factor of 13, and acts in the opposite direction (Ingold & Nathan 1936; Evans, Gordon & Watson 1937). In these reactions, however, the polar groups in the *para* position do not undergo chemical change, but exert an influence across the ring on the interaction of the hydroxyl ion with the polar group at its other end. The remoteness of the substituted group accounts for the smallness of the effect. The term $(1-LT)$ in such a case, though not known exactly, must have a positive value lying between unity, for *vacuo*, and 0·74 for benzene. Equation (24), with a constant value of E_n, is thus applicable, as has been shown in various ways (Moelwyn-Hughes 1936; Jenkins 1939).

AN ALTERNATIVE MECHANISM OF ACTIVATION

The general dynamical interpretation of the energy of activation (Heitler & London 1927) has been, after considerable and necessary modification, applied to reactions between ions and polar molecules in solution (Ogg & Polanyi 1935). The dominant idea is that, after removing a solvent molecule from the way, making an appropriate adjustment in energy, the critical increment of energy is chiefly expended in overcoming the strong repulsion exerted between the ion and the molecule at the close distances required for chemical change. This also is the idea which, in a more direct, but possibly less accurate, manner has been employed in the present treatment.

Work done recently in this laboratory in collaboration with Mr R. Hurst has suggested the possibility of another interpretation, which is that the energy of activation for certain reactions between ions and polar molecules in solution consists principally in the energy required to remove a solvent molecule from the fairly firm shell formed around the ion. If the critical increment of energy were exclusively that needed in this partial desolvation, it could be argued that the energy of activation of all reactant molecules with a common ion should have the same value, which is contrary to experience. A similar process of liberation from the solvent molecule thus seems necessary, though probably on a smaller scale, for the solute molecule. Two methods suggest themselves for pursuing the idea. The process of activation can be dynamically formulated as the withdrawal of a solvent molecule, in the field exerted by the solute molecule, from contact with the ion. Alternatively, statistical expressions can be derived for the fractional numbers of solute particles, both ions and molecules, which have neighbouring sites unoccupied by solvent molecules, and which are, on account of their incomplete solvation, more disposed to chemical reaction.

REFERENCES

Bergmann, E., Polanyi, M. & Szabo, A. 1933 *Z. phys. Chem.* B, **20**, 161.
Conrad, M. & Brückner, C. 1889 *Z. phys. Chem.* **4**, p. 631.
Evans, D. P., Gordon, J. J. & Watson, H. B. 1937 *J. Chem. Soc.* p. 1430.
Grant, G. H. & Hinshelwood, C. N. 1933 *J. Chem. Soc.* p. 258.
Heitler, W. & London, F. 1927 *Z. Phys.* **44**, 457.
Ingold, C. K. & Nathan, W. S. 1936 *J. Chem. Soc.* p. 222.
Jenkins, H. O. 1939 *J. Chem. Soc.* p. 1780.
Latimer, W. M., Pitzer, K. S. & Slansky, C. M. 1939 *J. Chem. Phys.* **7**, 108.
Lennard-Jones, J. E. 1936 In Fowler's *Statistical mechanics*, 2nd ed., pp. 306, 326. Cambridge University Press.
Moelwyn-Hughes, E. A. 1936 *Proc. Roy. Soc.* A, **157**, 667.
Moelwyn-Hughes, E. A. 1938*a* *Proc. Roy. Soc.* A, **164**, 295.
Moelwyn-Hughes, E. A. 1938*b* *J. Chem. Soc.* p. 779.
Moelwyn-Hughes, E. A. 1941 *Trans. Faraday Soc.* **37**, 279.
Moelwyn-Hughes, E. A. 1947 *The kinetics of reactions in solution*, 2nd ed., p. 227. Oxford University Press.
Nathan, W. S. & Watson, H. B. 1933 *J. Chem. Soc.* p. 890.
Ogg, R. A. & Polanyi, M. 1935 *Trans. Faraday Soc.* **31**, 604.
Slator, A. 1904 *J. Chem. Soc.* **85**, 1286.

The spectral response of a phosphor to infra-red radiation

By E. F. Daly,* *Department of Colloid Science, University of Cambridge*

(*Communicated by F. G. W. Roughton, F.R.S.—Received* 16 *August* 1948)

[Plate 12]

A determination of the infra-red spectral response of a copper-activated zinc cadmium sulphide phosphor at normal and at liquid air temperatures is described. Peaks in sensitivity at wave-lengths of 2·1 and 2·7μ are observed directly for the first time and are correlated with glow-curve data. The incidence of room temperature radiation on a phosphor at liquid air temperature is shown to make the observation of sensitivity beyond 3μ impracticable.

INTRODUCTION

The effect of infra-red radiation in stimulating or quenching the emission of light from phosphors has been known for a considerable time and has been made use of by many workers in detecting infra-red radiation of wave-lengths between the visible and a limit of 1·7 or 1·8μ (Bünger & Flechsig 1931; Lewschin, Antonow-Romanowsky & Tumerman 1934; O'Brien 1946; Paul 1946). An examination of the earlier results suggested that phosphors might provide means of detecting infra-red radiation of longer wave-length if suitable experimental conditions could be found, and might then be of some practical value in detecting hot objects or in observing absorption spectra. The use of phosphors at low temperatures to detect infra-red radiation in the 2 to 10μ region has in fact been suggested by Randall & Wilkins (1945). In order to explore this possibility some experiments were made, early in 1940, on the infra-red sensitivity of phosphors, in particular on that of a copper-activated zinc cadmium sulphide, G.E.C. luminescent powder $F(P)$. It was found that for this phosphor the stimulation of luminescence by infra-red radiation had a limit at 1·8μ when the specimen was at room temperature, but that the effect could be detected to a limit of 2·8μ when the specimen was cooled to liquid air temperature. Since observations on spectral sensitivity beyond 2μ do not appear to have been described previously in the literature, a brief account of the experiments is given below.

EXPERIMENTAL

The apparatus used in this work consisted of a cell in which a specimen of phosphorescent powder could be irradiated while at a low temperature, a mercury arc lamp for exciting luminescence, an infra-red source with either a monochromator or a spectrograph arrangement for projecting infra-red radiation on the phosphor specimen and a camera for recording the luminescence of the specimen under various conditions. The layout of the apparatus using a monochromator was as shown in figure 1, the spectrograph arrangement being similar but with the omission of the Wadsworth mirror and final condensing mirror and with the phosphor specimen at the focus normally occupied by the monochromator exit slit.

---

* I.C.I. Research Fellow.

A diagram of the phosphor cell is given in figure 2. The central pyrex tube, which had a short side-arm near the bottom, could be filled with liquid air for cooling. The edges of this side-arm were ground flat and a thin glass cover-slip cemented over the opening. A sheet of insulating material about 1 mm. thick and having a rectangular hole cut in it was cemented against the outer surface of the cover-slip and the hole filled with a layer of phosphor powder bound with a little turpentine. Turpentine was used because on evaporation it leaves little residue which may absorb infra-red radiation. The resulting 1 mm. thick layer of phosphor was in close

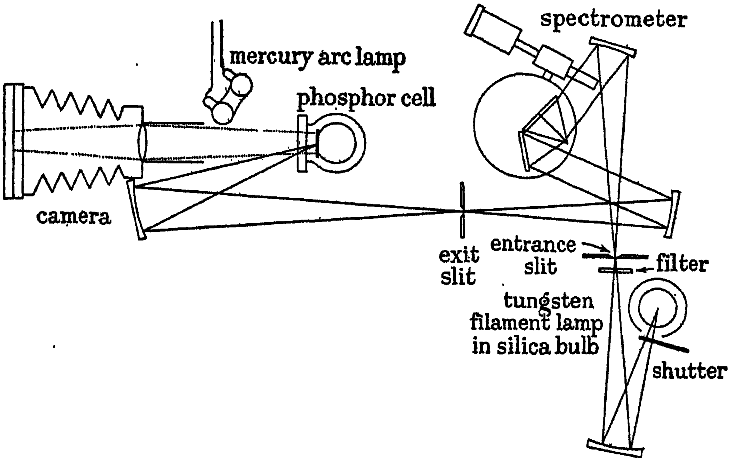


FIGURE 1. Monochromator for determining the infra-red spectral response of phosphors.

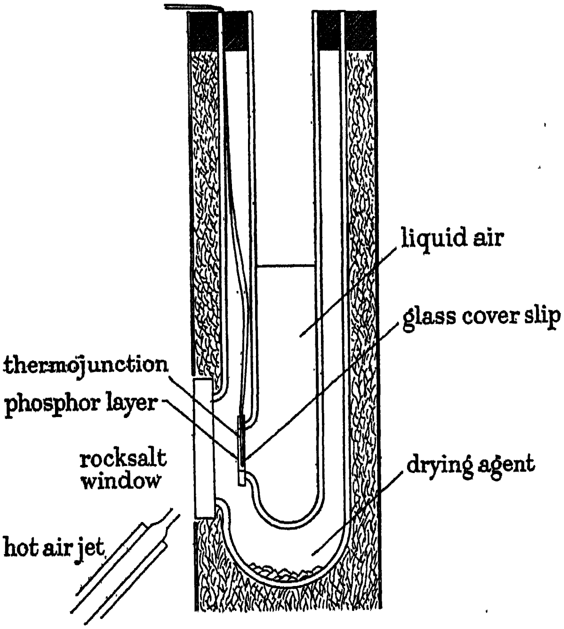


FIGURE 2. Cell for observation of phosphor layers at low temperatures.

contact with the middle of the cover-slip and thus with the cooling liquid and had embedded in it a thermo-junction of fine gauge copper and constantan wire used for measuring the temperature of the specimen. A second tube surrounded the inner tube to protect it from condensation of moisture at low temperatures and contained a small quantity of drying agent. This outer tube also had a side-arm closed with either a rocksalt or a cover-slip window located opposite the phosphor sample, the window being kept free from condensed moisture when in use by means of a jet of warm air directed against its outer surface.

Short wave-length light for exciting luminescence was provided by a small mercury arc in a silica envelope used in a position which was not varied with respect to the phosphor surface. A tungsten filament lamp in a silica bulb was used as the source of infra-red radiation and was run at a temperature such that its radiation maximum occurred at about $1\cdot2\mu$.

For measurements on the variation with wave-length of sensitivity of the phosphor towards infra-red radiation a technique was adopted by which the effects of both stimulation of luminescence and quenching were added but by which a fairly high photographic contrast was obtainable between irradiated and normal parts of the specimen. In experiments using the monochromator system the phosphor was first illuminated for 30 sec. by mercury arc light, and an infra-red image of the spectrometer exit slit then projected on to the surface of the phosphor for 25 sec. Five seconds later a photographic exposure lasting 10 sec. was made of the phosphor layer. The appearance of the phosphor surface at this stage was that of a rectangle of normal phosphorescent brightness crossed by a darker line corresponding to the part exposed to the infra-red image. The effectiveness of the infra-red radiation was measured by taking a microphotometer record of the plate blackening in the photograph of the phosphor surface, traversing the plate at right angles to the direction of the infra-red slit image, and obtaining from this record a quantity, q, roughly proportional to the percentage change in luminescence on irradiation. If b represents the light intensity transmitted by the photographic plate at the centre of the infra-red slit image, and a and c the intensities corresponding to the background just clear of this image on either side, $q = [b-\frac{1}{2}(a+c)]/\frac{1}{2}(a+c)$. Here q is independent of the absolute values of a, b and c provided that these fall within the range over which the optical density of the plate varies as the logarithm of the light intensity during exposure.

Since a single monochromator system was used adequate precautions had to be taken against scattered radiation present as impurity in the output. This was done by observing and comparing the spectra obtained when a number of infra-red filters of different transmission characteristics were used in turn in the beam. Three filters were found useful: A, a film of bitumen between thin glass cover-slips, which transmits infra-red radiation well from $1\cdot4$ to beyond 3μ; B, a combination of the bitumen filter with a 2 mm. cell of water, which has satisfactory transmission between $1\cdot2$ and $1\cdot8\mu$ but is opaque beyond 2μ; and C, black mica, which is opaque to $1\cdot4\mu$ radiation but transmits well between 2 and 3μ. Infra-red radiation from the tungsten lamp which was transmitted by each of these filters is plotted against wave-length in figure 3.

A small Hilger infra-red spectrometer with a quartz prism was used as the monochromator and exposures of phosphor luminescence, using an F 2·5 camera and panchromatic plates, were taken at spectral intervals of 0·2μ over the range 1 to 3μ. The mean width of the spectral band passed by the monochromator was 0·1μ over this range. Plots of q against wave-length for the phosphor used are shown for room temperature in figure 4 and for liquid air temperature in figure 5. The normal background phosphorescence is much weaker at liquid air temperature than at room temperature, and values of q from figures 4 and 5 are not strictly comparable as absolute measures of phosphor sensitivity.

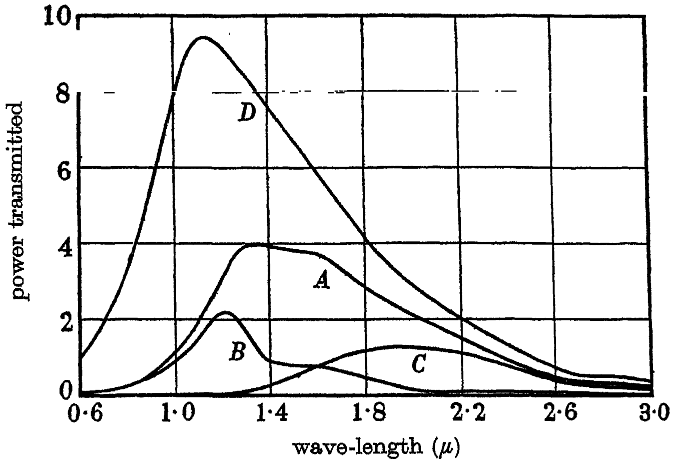


FIGURE 3. Tungsten lamp radiation transmitted by various filters. *A*, bitumen filter; *B*, bitumen + 2 mm. of water; *C*, black mica filter; *D*, full tungsten lamp radiation.

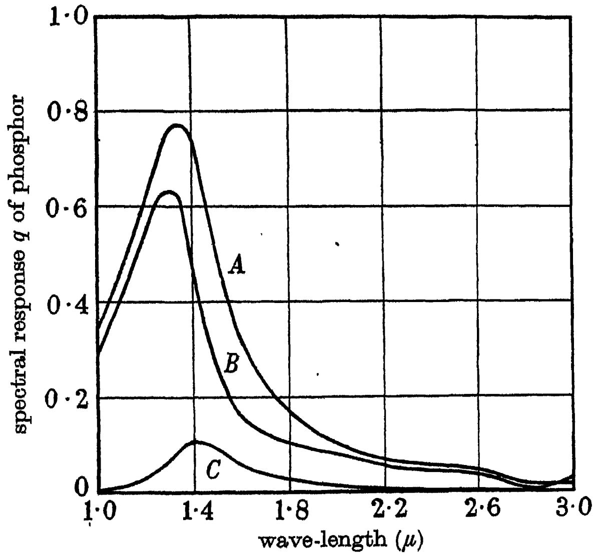


FIGURE 4. Spectral response of phosphor to infra-red irradiation at room temperature. *A*, bitumen filter in beam; *B*, bitumen + water filter in beam; *C*, black mica filter in beam.

At both liquid air temperature and room temperature the sensitivity of the phosphor to infra-red radiation near 1·4μ is very high. Since the infra-red source has a maximum radiation intensity near 1·2μ it is to be expected that sufficient impurity of this active 1·4μ radiation will be present in the monochromator output at longer wave-lengths to cause spurious effects. Thus in the room temperature sensitivity curve for the phosphor, using the bitumen filter (figure 4*A*), the 1·4μ band appears to have a tail extending to 3μ, while in the corresponding experiment using the black mica filter, which severely restricts the passage of 1·4μ radiation (figure 4*C*), there is no tail although the filter readily transmits wave-lengths in the 2 to 3μ region.

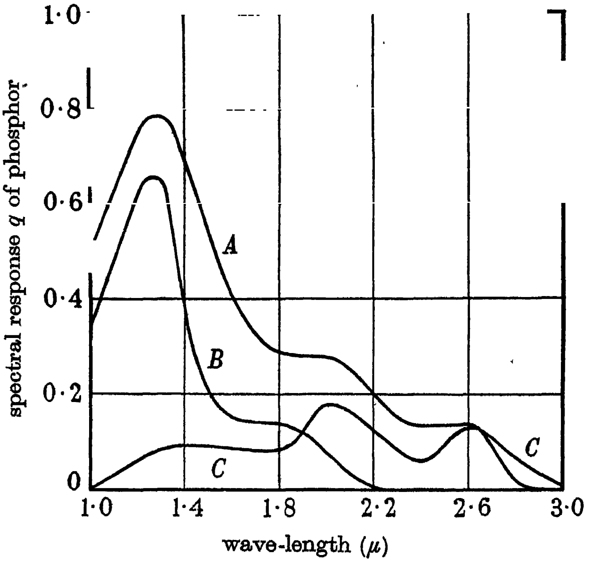


FIGURE 5. Spectral response of phosphor to infra-red irradiation at liquid air temperature. Filters *A*, *B* and *C* as in figure 4.

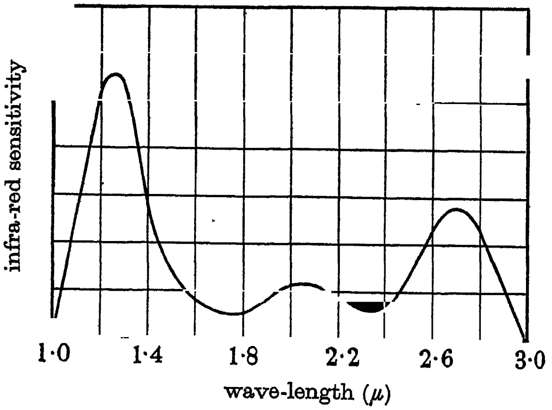


FIGURE 6. Approximate infra-red sensitivity curve of phosphor at liquid air temperature for uniform spectral energy distribution.

At liquid air temperature subsidiary maxima in phosphor sensitivity are observed near 2·1 and 2·7μ, though owing to the much lower available infra-red intensity the effect is not nearly so obvious as near 1·4μ. From the low temperature sensitivity curves using either black mica or bitumen filters (figure 5*A* and *C*) the approximate wave-lengths of the two subsidiary maxima may be determined. An approximation to the actual variation in efficiency of the infra-red radiation with wave-length is made in figure 6. This curve is based on the observed plot of q against wave-length for the case of the black mica filter (figure 5*C*), correction being made for filter transmission and source intensity variation by dividing the values of q by the corresponding transmitted intensity values given in figure 3.

In table 1 wave-lengths of the three infra-red sensitivity maxima are given, together with corresponding energy transition values in eV.

TABLE 1

positions of infra-red sensitivity maxima in μ	energy differences of corresponding transitions (eV)
1·26	0·985
2·08	0·60
2·70	0·46

When the sensitivity of the copper-activated zinc cadmium sulphide phosphor to infra-red radiation of wave-length as great as 2·8μ had been established, an attempt was made to photograph absorption spectra in this region using the spectrograph arrangement mentioned above, with lengthened times of infra-red irradiation and photographic exposure. Spectra obtained in this way are illustrated in figure 8, plate 12, in which the absorption of a water film near 2μ and in the 2·7μ region may be detected. The method yields spectra which are not comparable in resolution with those obtained by conventional techniques, but it offers an alternative, over a limited spectral range, to the method described by Czerny & Mollet (1937) in which the infra-red radiation, by partially evaporating a liquid film, gives rise to spectral patterns.

DISCUSSION OF RESULTS

The explanation generally accepted of the mechanism of infra-red interaction with phosphors involves the presence of electron traps, presumably associated with impurities or imperfections in the crystal lattice, from which electrons may be released by thermal agitation in the crystal or by infra-red radiation. It is found in the case of the copper-activated zinc cadmium sulphide phosphor used in the above experiments that the predominant effect on irradiation by infra-red of an excited specimen is a brightening of the luminescence. It has also been observed during the present work that the emission spectrum of the normal luminescence of this phosphor, which is characteristic of the copper activator, appears to be the same whether the phosphor is emitting at room temperature or at liquid air temperature. These two facts would suggest that, whatever the electron trapping level concerned, the principal mechanism by which an electron released from a trap reaches the ground state involves the same luminescent transition. Thus the presence of structure in the infra-red sensitivity curve indicates that the electron trapping levels have some characteristic distribution in depth, being concentrated about the values 0·985, 0·60 and 0·46 eV below some arbitrary level from which the luminescent transition to the ground state may commence.

A similar conclusion has been reached by Randall & Wilkins (1945) on the basis of the glow curves which they have observed for similar phosphor specimens, and attribute to the progressive release of more and more deeply trapped electrons with rise in temperature of the phosphor specimen. It is therefore of interest to note that in figure 7 there is considerable resemblance between the pattern of trap distribution for a copper-activated zinc cadmium sulphide phosphor observed by infra-red methods and the corresponding distribution deduced by the simplified method suggested by Randall & Wilkins from the glow curve for one of their copper-activated zinc sulphide specimens.

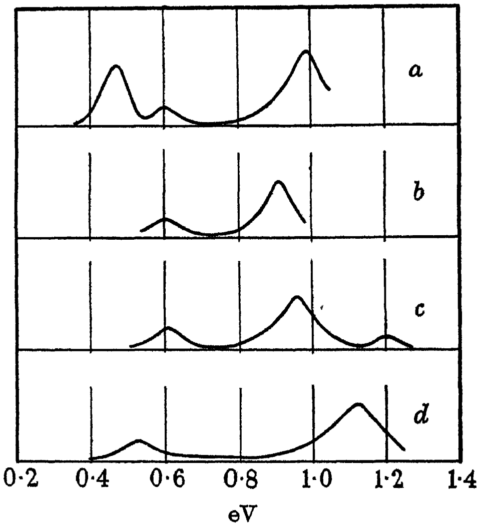


FIGURE 7. Approximate distribution of traps according to depth in eV. (Diagrammatic.) *a*, distribution from infra-red sensitivity spectrum, present work; *b*, distribution from approximate glow-curve data, present work; *c*, distribution calculated from glow-curve data of Randall & Wilkins for a copper-activated zinc sulphide phosphor; *d*, distribution calculated from glow-curve data of Wilkins & Garlick for a copper-activated zinc cadmium sulphide phosphor.

It may be noted that during warming of a phosphor layer which had been excited by mercury arc radiation while at liquid air temperature an appreciable increase in the rate of rise of temperature occurred near the point of maximum luminescence of the glow curve. This effect, while not further investigated, appeared large enough to make calorimetric measurements possible and might provide further information on the luminescence mechanism.

It seems fairly clear from the experiments described above and from the results of other workers (Randall & Wilkins 1945; O'Brien 1946) that at room temperature

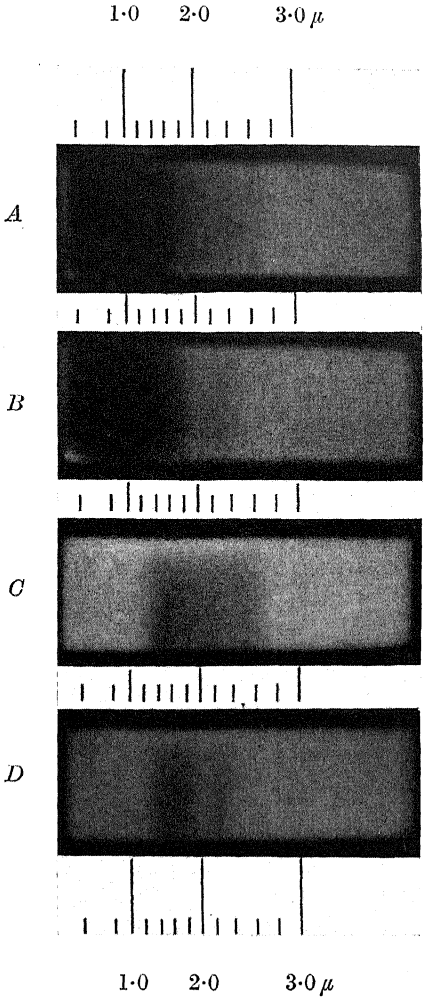


Figure 8. Photographs of phosphorescent zinc cadmium sulphide layer at liquid air temperature showing darkening caused by infra-red spectra projected on to the layer. *A*, no filter in infra-red beam; *B*, 0·5 mm. layer of water in infra-red beam; *C*, black mica filter in beam; *D*, black mica filter + 0·5 mm. layer of water in beam.

thermal agitation in the phosphor crystals will not permit electrons to remain trapped when the trap depth is less than about 0·6 eV. If the liberation of electrons from traps is assumed to be proportional to $\exp(-E/kT)$, then the trap depth, E, necessary to retain electrons when the temperature, T, is reduced from 300 to 100° K would be reduced from 0·6 to 0·2 eV. Therefore, if the room temperature sensitivity threshold is at 2μ (0·6 eV) the threshold at liquid air temperature ought to be in the region of 6μ (0·2 eV). In fact no infra-red sensitivity has been observed in the phosphor between 2·8 and $3{\cdot}5\mu$, the long wave-length limit of quartz prism transmission.

Since there does not appear to be any reason why shallow electron traps (0·2 to 0·4 eV) should not occur in the phosphor investigated, it is of interest to consider the significance of the $2{\cdot}8\mu$ limit. Precise experimental determination of this threshold is made difficult by the presence of quartz absorption and by the low infra-red radiation intensities available near 3μ, but it will be shown on theoretical grounds that for the experimental conditions used the threshold imposed by the incidence of room temperature black-body radiation on the phosphor would not be greatly different from the observed value of $2{\cdot}8\mu$.

From measurements by Urbach, Pearlman & Hemmendinger (1946) and by Ellickson (1946) on the total amount of light which a phosphor can store and on the quantum efficiency of infra-red stimulation of luminescence it may be assumed that the incidence of some 10^{18} quanta/cm.^{3} of phosphor, at a wave-length shorter than the threshold, would liberate practically all the trapped electrons. For the liberation of trapped electrons in a phosphor layer 1 cm.^{2} in area and 0·01 cm. thick, in a time of 100 sec., 10^{14} quanta/sec./cm.^{2} are required. A black body at 300° K will radiate $0{\cdot}7 \times 10^{14}$ quanta/sec./cm.^{2} at wave-lengths shorter than 3μ, $1{\cdot}8 \times 10^{15}$ quanta/sec./cm.^{2} at wave-lengths shorter than 4μ and $1{\cdot}3 \times 10^{16}$ quanta/sec./cm.^{2} at wave-lengths shorter than 5μ. It is therefore clear that there is little hope of observing sensitivity towards radiation of wave-lengths between 3 and 6μ in a phosphor cooled to 100° K unless extreme precautions are taken to exclude room temperature radiation, which would be expected to liberate within a short time electrons in traps shallower than 0·4 eV. It may thus be assumed, until measurements in the absence of room radiation have been carried out, that the $2{\cdot}8\mu$ threshold obtained in the present work is a consequence of the experimental conditions rather than a property of the phosphor.

In conclusion it may be said that the above discussion indicates the experimental limitation to be overcome in searching for longer wave-length infra-red sensitivity of phosphors, but also suggests that such a search, carried out by measurements on either luminescence or photoconductivity, is likely to lead to interesting results. Further investigation of long wave-length infra-red sensitivity in phosphors might help in the elucidation of the similar but more complex properties of semiconductors such as lead sulphide or lead telluride.

The experimental work was carried out during the tenure of an 1851 Exhibition Science Scholarship, the award of which the author gratefully acknowledges. He wishes particularly to express his indebtedness to Dr G. B. B. M. Sutherland, who suggested the problem, for his interest and encouragement.

REFERENCES

Bünger, W. & Flechsig, W. 1931 *Z. Phys.* **69**, 637.
Czerny, M. & Mollet, P. 1937 *Z. Phys.* **108**, 85.
Ellickson, R. T. 1946 *J. Opt. Soc. Amer.* **36**, 264, 501.
Garlick, G. F. J. & Wilkins, M. H. F. 1945 *Proc. Roy. Soc.* A, **184**, 408.
Lewschin, W. L., Antonow-Romanowsky, W. W. & Tumerman, L. A. 1934 *Phys. Z. Sowjetun.* **5**, 611.
O'Brien, B. 1946 *J. Opt. Soc. Amer.* **36**, 369.
Paul, F. W. 1946 *J. Opt. Soc. Amer.* **36**, 175.
Randall, J. T. & Wilkins, M. H. F. 1945 *Proc. Roy. Soc.* A, **184**, 366.
Urbach, F., Pearlman, D. & Hemmendinger, H. 1946 *J. Opt. Soc. Amer.* **36**, 372.

The emission of the negative system of nitrogen from the upper atmosphere and the significance of the twilight flash in the theory of the ionosphere

By D. R. Bates, *Department of Mathematics, University College, London*

(*Communicated by H. S. W. Massey, F.R.S.—Received* 23 *September* 1948
—*Read* 3 *March* 1949)

The cause of the emission of the negative band system of nitrogen from the upper atmosphere during twilight is investigated. A study is made of the two possible excitation mechanisms,

$$N_2(X\ {}^1\Sigma_g^+) + h\nu \rightarrow N_2^+(B\ {}^2\Sigma_u^+) + e$$

and

$$N_2^+(X\ {}^2\Sigma_g^+) + h\nu \rightarrow N_2^+(B\ {}^2\Sigma_u^+).$$

It is shown that the latter is far more effective than the former, irrespective of the assumptions adopted regarding the solar flux in the unobservable spectral region. From the transition probability associated with it (which is evaluated in the appendix) combined with various intensity estimates, an upper limit is obtained for the number of N_2^+ ions normally present in the E and F layers during twilight. It appears that N_2^+ ions form but a minute fraction of the total ion content. The significance of this in the theory of the formation of the ionized layers is discussed. The simplest interpretation is that ionization of molecular nitrogen is unimportant; and a reasonable scheme that invokes only the ionization of oxygen atoms and molecules is available. However, by introducing certain arbitrary assumptions a more elaborate interpretation is conceivable so that the view that the E layer arises from the action of high-energy coronal photons, which ionize all atmospheric constituents, cannot be finally rejected. Various aspects of the layers are discussed, and observational and experimental work, which might yield evidence on the ionization mechanisms operative, is suggested.

It is pointed out that the remarkable rarity of N_2^+ ions proves conclusively that recombination between the charged particles present in the ionosphere cannot be the origin of the nocturnal radiation of the nitrogen band systems. On some occasions the resonance emission at twilight is of unusually high intensity. It is presumed that this is due to incident charged particles increasing the concentration of N_2^+ ions. The possible contribution that these charged particles may make to the night-sky light by direct excitation collisions is briefly examined.

Sunlit aurorae (which are essentially similar to the twilight flash) are also discussed.

1. INTRODUCTION

The negative system of nitrogen is emitted by the molecular ion in a radiative transition from its first excited state $B\,^2\Sigma_u^+$ to its ground state $X\,^2\Sigma_g^+$. $B\,^2\Sigma_u^+$ lies 3·15 eV above $X\,^2\Sigma_g^+$ and 18·7 eV above the ground state of the neutral molecule $X\,^1\Sigma_g^+$. The bands are situated in the violet end of the spectrum.

The sunlit and normal aurorae are, of course, the most arresting light source in the upper atmosphere, and in the spectra of these the bands of the negative system are very prominent as the work of Vegard (1939) and others has shown conclusively. It is possible that they also form a weak and erratic component of the night-sky emission. The relevant wave-length region has been carefully studied by several groups of workers, but the large number of radiations occurring, together with the unavoidable lack of precision in the measurements, make the certain identification of any system difficult as they lead to the possibility of fortuitous coincidences. In an analysis of the problem Dufay & Déjardin (1946) point out that some fourteen negative bands and some forty-nine night-sky bands are located between λ 3086 and λ 3935, but that between these there is only agreement within 1 Å in two instances—which is just what would be expected if chance alone were responsible. This shows that the system is certainly not highly developed. However, if the excitation process were the collision between fast electrons and neutral nitrogen molecules high development would not occur. Indeed, in a feeble source probably the only members of the system that would be observed would be the (0, 0) band λ 3914 and (with about one-third to one-quarter the intensity) the (0, 1) band λ 4278. Irregular night-sky emissions have been reported close to both of these, the former being the better established (cf. Elvey, Swings & Linke 1941, λ 3914; Dufay 1943*a*, λ 3914 and λ 4278). Certain aspects of the general nocturnal luminescence suggest that for it, in addition to the main excitation processes, another, of the collision type mentioned, may possibly have to be invoked giving essentially a faint non-polar aurora (cf. Bates 1948).* In this case complete absence of the negative system would scarcely be anticipated. Further work on the final establishment of the identification of the two emissions referred to, and a study of their variations, and how these are correlated with other effects, would be of value.

The situation during twilight (which is the principal subject of the present discussion) is very different. At first the observational evidence seemed contradictory. Slipher (1933) and later Elvey (1942), in the United States, found that the negative system appeared strongly at sunrise and sunset; but Gauzit (1938) in France, and Cario (1938) in Germany, failed to detect the phenomena. Dufay & Dufay (1947) in an important recent paper have elucidated the reason for the apparent discrepancy. Working in France they found that the so-called flash varies greatly in intensity; on some occasions it is too weak to be observed, and on others it attains considerable prominence. They established too a positive correlation between the intensity and the magnetic activity and (what is probably of more direct significance) between the intensity and the occurrence of aurorae in Norway. The possibility that the variation might not be real, but might instead arise from some effect such as changes in the atmospheric transmissibility, would appear to be excluded by these correlations.

---

* See also § 5.

2. GENERAL REMARKS ON THE EXCITATION PROCESS YIELDING THE TWILIGHT FLASH

Two different processes have been put forward for the provision of the excited nitrogen molecular ions responsible for the twilight flash. Saha (1937) has strongly advocated the view that they arise through the action of solar radiation on neutral nitrogen molecules,

$$N_2(X\,{}^1\Sigma_g^+) + h\nu \rightarrow N_2^+(B\,{}^2\Sigma_u^+) + e. \tag{1}$$

This hypothesis appears to have been widely accepted by astrophysicists (cf. Hunter 1942). Wulf & Deming (1938) have, however, suggested that the process involved is simply one of direct resonance,

$$N_2^+(X\,{}^2\Sigma_g^+) + h\nu \rightarrow N_2^+(B\,{}^2\Sigma_u^+). \tag{2}$$

It is clearly important to decide which of these alternatives is correct.

Only photons of energy above 18·7 eV can give rise to process (1). The flux of these from the conventional representation of the sun as a black body at 6000° K is much too small, as was fully realized by Saha, who postulated that at such high energies the actual emission is many orders greater than that from the black-body model. Recent work does not favour this (cf. Hoyle 1949), but in the present state of knowledge of solar physics it probably cannot be excluded; and, indeed, Johnson (1947) has discussed a possible mechanism. However, it is now known from the theory of the production of the F_1 layer (cf. Bates & Massey 1946) that at, and some way beyond, the first ionization potential of atomic oxygen, 13·55 eV, the flux is *not* greatly in excess. Though it might perhaps have been accepted that the solar emission is extremely intense in the high-energy region, it is much less plausible that it approximates to that from the black-body model near 13·55 eV (which almost coincides with the Lyman limit 13·53 eV) but rises to be vastly greater than it in the neighbourhood of 18·7 eV. Apart from this, molecular nitrogen ions are of necessity formed at a rate at least equal to the rate of photon emission through (1). It will be shown later (§ 3·4) that this rate alone is sufficient to ensure their presence in sufficient concentration for process (2) to be predominant; and reasons will be given for supposing that a more effective ionizing mechanism is in operation.

2·1. Other aspects of the phenomenon will now be discussed on the two hypotheses.

(i) Dufay & Dufay observed that the flash is of short duration. They were rarely able to detect any trace of it when the sunlit region lay above 130 km., and they found that the main effect is limited to the period when the atmosphere in a layer some 50 km. thick centred at about 100 km. is illuminated. This gives directly the approximate altitude of the emission if it is process (2) that is responsible; but if it is process (1) the altitude is much greater since the earth's effective shadow radius is increased because of the rapid attenuation of the active solar radiation by the atmospheric gases. The relevant absorption cross-section of atomic oxygen is approximately $1{\cdot}6 \times 10^{-17}$ cm.^{2} and that of molecular nitrogen is probably at least as great (cf. Bates & Massey 1946). As a consequence even at vertical incidence there can be no appreciable penetration below about 300 km. When twilight is at the stage that low-energy photons are just reaching the 100 km. level, high-energy (18·7 eV) photons are perhaps only reaching the 500 km. level; thus it is from here that any emission of the negative bands that result from process (1) must arise. Judging from the evidence provided by the apparent great scale height of the F_2 layer (Appleton 1937), the luminosity curve of aurorae (Harang 1945) and the variation of sunrise enhancement of the forbidden red line of oxygen (Elvey & Farnsworth 1942), one would scarcely expect that if the flash originates at such an altitude it would be as brief as is observed. With process (2) there is no such difficulty.

(ii) The great variability of the intensity discovered by Dufay & Dufay must presumably, on Saha's hypothesis, be attributed to fluctuations in the solar emission beyond 18·7 eV. Now the rate of production of $N_2^+(B\,^2\Sigma_u^+)$ necessary to explain the flash (*vide infra*) would lead to a striking increase in the electron density in the F_2 layer; furthermore, unless the changeable spectral region did not extend down to near 13·55 eV the F_1 layer would also be augmented. Such effects do not appear to have been recorded; admittedly, however, the associated magnetic storms may confuse the issue. The correlation with auroral displays in Norway is of considerable interest. It would seem to require that the enhancement of the high-energy solar radiation and the emission of the corpuscles responsible for the aurorae are connected. However, the commonly accepted one-day time difference between the arrival of radiation and corpuscles is apparently absent.

On Wulf & Deming's hypothesis the lack of regularity of the flash must be due to changes in the number of N_2^+ ions present, as, of course, the solar radiation of the comparatively low energy (3·15 eV) required for resonance excitation is essentially of constant intensity. One obvious possibility is that the change is brought about by additional ionization being produced near the 100 km. level by charged particles. It may be remarked that Appleton & Naismith (1940) have found some evidence for such an effect in their work on the sporadic E layer. Papalexi (1946), too, claims that measurements during eclipses indicate that solar corpuscles influence all the main ionospheric layers. The matter will be discussed in more detail below, and it will be shown that a weak source of ionization is sufficient.

In connexion with the two possible causes of the variability it is clearly desirable that the dependence of the flash on latitude be studied; observations near the equator should be particularly revealing.

(iii) Dufay & Dufay commented that the relative intensity of the different bands of the negative system as emitted at twilight is peculiar in that there is a greater development of those bands arising from excited vibrational levels than is usual in aurorae. Calculations have been performed on the intensity distributions that might arise from the two mechanisms under consideration. Details have been published in another paper (Bates 1949). It is sufficient here to note that qualitatively at least the observed distribution is consistent with either mechanism; but with (1) it is necessary to assume that the neutral nitrogen molecules in the high atmosphere are themselves vibrationally excited to an extent corresponding to some thousands of degrees absolute, whereas with (2) no additional postulate need be made. Observational work on the rotational structure of the bands would be valuable.

3. THE ABSOLUTE INTENSITY

3·1. *The Saha hypothesis*

Direct evaluation of the absolute intensity resulting from process (1) is not at present feasible because of lack of basic data on, for example, the solar flux in the relevant spectral region. However, an *upper limit* to the intensity can be set by using information obtained from the study of the ionosphere. As mentioned before the incident radiation concerned will be absorbed high in the atmosphere. It will tend to form a layer; and the most favourable assumption is that it forms the F_2 layer. Clearly the photon emission rate that results cannot be greater than the ion production rate.* This last can be found, at least approximately, from the radio measurements.

What is actually required is the total rate of production of ions at twilight in a column of unit cross-section lying along the line of observation. The calculation is not difficult.

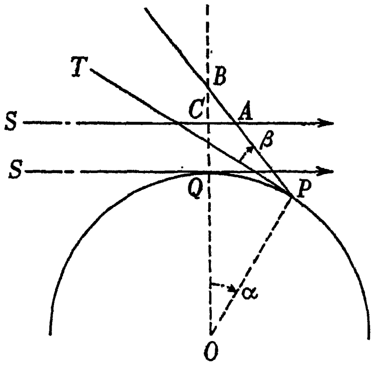


FIGURE 1

Referring to figure 1, P is the observer, S the sun, SQ is a ray just touching the earth's surface (neglecting refraction), SCA is a representative other ray, $OQCB$ is the perpendicular to these rays from O the earth's centre, PT is the tangent line at P, and PAB is the line of observation. The following notation is used:

Angle of depression of sun $P\hat{O}Q = \alpha$.
Angle of elevation of line of observation $T\hat{P}B = \beta$.
Radius of earth $= R$ km.
Altitudes of A, B and $C = h_A$, h_B and h_C km. respectively ($\ll R$).
Altitude of noon level of maximum absorption $= h_m$ km. ($\ll R$).
Nitrogen particle densities/cm.^{3} at altitudes h_A, h_C and $h_m = n_A(\mathrm{N}_2)$, $n_C(\mathrm{N}_2)$ and $n_m(\mathrm{N}_2)$ respectively.
Scale height in region concerned $= H$ km.
Noon ion production rate at level of maximum absorption q_m/cm.^{3}/sec.

---

* It is, in fact, less, as atomic oxygen *must* give a considerable contribution (*vide infra*).

It can readily be seen that

$$h_A = h_C + \{(h_B - h_C)^2/2(R+h_C)\}\cot^2(\alpha+\beta), \tag{3}$$

$$h_B = R\{\sec(\alpha+\beta)\cos(\beta) - 1\}, \tag{4}$$

$$n_A(\mathrm{N}_2) = n_m(\mathrm{N}_2)\exp\{-(h_A - h_m)/H\}, \tag{5}$$

$$n_C(\mathrm{N}_2) = n_m(\mathrm{N}_2)\exp\{-(h_C - h_m)/H\}. \tag{6}$$

Elementary analysis shows that the number of nitrogen molecules traversed by the representative ray before reaching A is given by

$$\mathcal{N}_A(\mathrm{N}_2) = n_C(\mathrm{N}_2)\{\tfrac{1}{2}\pi H(R+h_C)\}^{\frac{1}{2}}\{1 \pm P(x)\}/\mathrm{cm.}^2, \tag{7}$$

the positive sign being taken if C is below B and the negative if C is above B; $P(x)$ is here the probability integral,

$$\frac{2}{\pi^{\frac{1}{2}}}\int_0^x \mathrm{e}^{-t^2}\,dt \tag{8}$$

and

$$x = \{h_B - h_C\}\{2H(R+h_C)\}^{-\frac{1}{2}}\cot(\alpha+\beta). \tag{9}$$

Since the attenuation of the solar radiation at A is

$$\exp\{-\mathcal{N}_A(\mathrm{N}_2)/Hn_m(\mathrm{N}_2)\}, \tag{10}$$

the ion production rate there is simply

$$q_m\{n_A(\mathrm{N}_2)/n_m(\mathrm{N}_2)\}\exp\{-(\mathcal{N}_A(\mathrm{N}_2) - Hn_m(\mathrm{N}_2))/Hn_m(\mathrm{N}_2)\}. \tag{11}$$

The necessary integration along the observing direction can be carried out numerically using the relations (3) to (11). Two ionospheric parameters H and q_m appear; the former was taken to be 100 km., which is probably not greatly in error (even during magnetic storms) for the region above the F_2 layer (cf. Harang 1945); the latter was taken to be 80/cm.^{3}/sec. (based on the figures adopted by Bates & Massey (1946) for their model ionosphere*). R is, of course, 6370 km. The computations were performed for $\alpha = 10, 15, 20°$ and $\beta = 10, 20°$. This covers the observational range.

For comparison with observation it is desirable to have the rate of emission of a single band (rather than of the whole band system) and to make allowance for atmospheric extinction. The strong (0-0) band λ 3914 seems most suitable for study. It may be assumed to contribute about half the full number of photons. The approximate transmissibilities can be calculated in the usual manner,† the extinction coefficient (used as an exponential factor) being 0·378/standard atmosphere (as found by interpolating between values at neighbouring wave-lengths given by Barbier 1947). Corrections can thus be applied to yield the band's effective rate of emission. The final results are given in table 1. Comment on the absolute magnitudes will be deferred until later. Attention may, however, be drawn to the gradualness of the diminution of the intensity as the altitude of the lowest illuminated level is increased. As remarked earlier there is apparently disagreement here between the predictions of Saha's hypothesis and the observational data (cf. § 2). The discrepancy is enhanced when allowance is made for the background continuum (due to Rayleigh scattering) at small angles of depression.

---

* Unless otherwise specified the data will be from this source.

† The simple first-order formula was used. Though admittedly this is far from accurate the errors caused are unlikely to affect the main argument—especially as the same formula is also used in calculations on the Wulf-Deming hypothesis (§ 3·2).

TABLE 1

solar depression α (°)	observing elevation β (°)	altitude of lowest illuminated level on line of observation (km.)	total rate of ion production/cm.^{2} column along line of observation ($\times 10^8$/sec.)	upper limit for effective rate of emission of 0-0 band/cm.^{2} column along line of observation ($\times 10^7$/sec.)
10	10	55	6·9	4·0
15	10	110	4·5	2·5
20	10	180	2·5	1·4
10	20	70	3·8	6·3
15	20	140	2·1	3·5
20	20	230	1·0	1·7

3·2. *The Wulf-Deming hypothesis*

If an atom, with a single excited state at energy $h\nu$ above the ground state, is bathed in black-body radiation of temperature T and dilution Ω, and if A is the spontaneous transition probability, then the rate at which it emits photons is clearly

$$\Omega A \exp(-h\nu/kT)/\text{sec.} \tag{12}$$

h and k are, of course, Planck's and Boltzmann's constants respectively. An equivalent expression, more convenient for the present purpose, is

$$\frac{c^3}{8\pi h\nu^3}\sigma(\nu)\,A/\text{sec.}, \tag{13}$$

where c is the velocity of light and $\sigma(\nu)$ is the energy of the radiation field in ergs/unit frequency range/cm.^{3}. The value of $c^3/8\pi h$ is $1{\cdot}62 \times 10^{56}$ cm.^{3}erg^{-1}sec.$^{-4}$. These formulae (with an additional factor allowing for the relative weights of the two states which is omitted here as it reduces to unity for the ${}^2\Sigma_u^+$-${}^2\Sigma_g^+$ transition under consideration) have frequently been used in discussions of twilight enhancement effects.

For the case of molecules some modification is necessary because of complications caused by the different vibrational levels. Both in absorption and in emission, changes in the vibrational quantum number may occur. As a result even if initially all the molecules were in the zeroth vibrational level they would become distributed among the other vibrational levels, an equilibrium distribution being rapidly reached in which the numbers entering balance those leaving. This equilibrium distribution depends, of course, on the values of ν, $\sigma(\nu)$ and A associated with individual bands of the system and also on the extent of collision deactivation. It has been investigated by Bates (1949). The results obtained are summarized in table 2 which gives for each vibrational level i, the relative population n_i. Two extreme conditions are represented: in the trivial case (1) deactivating collisions predominate and in case (2) they are absent—the former distribution is best used at low altitudes (200 km. or less) and the latter at great altitudes (400 km. or more).* For simplicity, levels with $i > 2$ are grouped together. The inaccuracy introduced is unlikely to be serious.

TABLE 2

i	0	1	2	remainder
n_i: case 1	1	0	0	0
case 2	0·55	0·25	0·13	0·07

The photon emission/illuminated ion can now be computed without difficulty. Using (13) it is obviously given by

$$P = \frac{c^3}{8\pi h}\sum_k\sum_i\{n_i\sigma(\nu_{ik})A_{ik}/\nu_{ik}^3\}/\text{sec.}, \tag{14}$$

the symbols being as before with the first subscript denoting the vibrational quantum number of the ground state and the second that of the excited state. A_{ik}/ν_{ik}^3 can be expressed as Gp_{ik}, where G is a constant of the whole band system and p_{ik} is a function of the vibrational quantum numbers (but not explicitly of the frequency) having the property

$$\sum_i p_{ik} = \sum_k p_{ik} = 1. \tag{15}$$

In the appendix G is shown to be about $3{\cdot}8 \times 10^{-38}$ sec.². The p's have been given by Bates (1949); they are quoted in table 3. There remains $\sigma(\nu_{ik})$ the energy densities. These were determined by integrating over the rotational bands, a weighting corresponding to a temperature of 0° C being assumed. For some of the bands the wavelengths were already available; for the others they were computed using the known molecular constants (Fassbender 1924; Coster & Brons 1932). The solar flux distribution curve was taken from the atlas published by Minnaert, Mulders & Houtgast (1940), the absolute scale in the various spectral regions being fixed by the measurements of Pettit (1940). By combining the results with those in table 2 and table 3 the values of $\sum_i n_i p_{ik}\sigma(\nu_{ik})$ shown in table 4 were obtained. Substitution in (14) then gives P, the photon emission/illuminated ion, to be 0·11/sec. (case 1) and 0·13/sec. (case 2).

TABLE 3. VALUES OF p_{ik}

	ground state			
vibrational quantum number	0	1	2	remainder
excited state 0	0·67	0·26	0·06	0·01
1	0·27	0·24	0·33	0·16
2	0·06	0·40	0·06	0·48
remainder	0·00	0·10	0·55	0·35

It is desirable as in § 3·1 to derive the rate for the emission of a single band (the 0-0 band being chosen) and to correct for atmospheric extinction. In addition, it is convenient to express the result in terms of the effective photon yield Q for an ion layer of thickness t km. and of mean ion concentration n/cm.³. Clearly,

$$Q = 10^5 tn\{\phi\chi\psi\}\,P/\text{cm.}^2\ \text{oblique column/sec.}, \tag{16}$$

---

* The altitudes quoted are rough estimates only.

ϕ, χ and ψ being, respectively, the fraction of the total photon emission in the chosen band, the transmissibility of the atmosphere to this band, and the ratio of the apparent thickness of the layer along the observing direction to the true thickness, and being given by the following expressions:

$$\phi = \{c^3G/(8\pi hP)\}\left\{p_{00}\nu_{00}^3\left(\sum_i n_i p_{i0}\sigma(\nu_{i0})\right)\right\}\left\{\sum_k p_{0k}\nu_{0k}^3\right\}^{-1}, \tag{17}$$

$$\chi = \exp(-A\,\mathrm{cosec}\,\beta), \tag{18}$$

$$\psi = \{1-(R/(R+z))^2\cos^2\beta\}^{-\frac{1}{2}}. \tag{19}$$

A is here the extinction coefficient/standard atmosphere (already used in § 3·1), z is the altitude of the ion layer, which may be taken to be 100 km., and the remaining symbols are as previously defined. Using the numerical data given above (in particular tables 2, 3 and 4), these expressions were evaluated and table 5, which contains in its last two columns the final results required, was compiled.

TABLE 4

k		0	1	2	remainder
$\sum_i n_i p_{ik}\sigma(\nu_{ik})$ (units 10^{-20} erg. cm.$^{-3}$ sec.):	case 1	$1{\cdot}5_2$	0·27	0·07	0·00
	case 2	$1{\cdot}1_2$	0·45	0·30	0·17

TABLE 5

	ϕ				Q/tn	
β	case 1	case 2	χ	ψ	case 1	case 2
10°	0·60	0·40	0·11	4·1	3100/sec.	2400/sec.
20°	0·60	0·40	0·33	2·6	5700	4500

Before leaving the problem of the intensity on the Wulf-Deming hypothesis it is necessary to investigate the passage through the air of the solar radiation causing the excitation. The equivalent path length that the representative ray SC (figure 1) traverses before encountering the ion layer can be seen from equation (7) to be

$$[n_C\{2\pi H(R+h_C)\}^{\frac{1}{2}}/n_G H_G] \text{ atmospheres,} \tag{20}$$

where n_G and H_G are respectively the particle concentration and scale height at ground-level under standard conditions, and the other symbols are as in § 3·1. This expression applies strictly only to an isothermal atmosphere, but if H is taken to be the local scale height at C, it represents a satisfactory approximation for the actual atmosphere, since the major contribution to the equivalent path arises from a relatively narrow altitude range just above the lowest point passed through by the ray. Knowing the equivalent path, the attenuation suffered can be determined at once. Calculations were performed for the three wave-lengths λλ 3500, 4000 and 4500. The atmospheric parameters adopted were based on the work of Whipple (1943), and the extinction coefficients (as before) on that of Barbier (1947). They are given, together with the attenuation derived, in table 6.

It can be seen that the atmosphere can crudely be regarded as increasing the earth's radius by about 20 km. At any instant during twilight the lowest altitude of the ion layer along the observing direction that is struck by the active solar radiation is greater than the lowest so-called illuminated altitude by about the same amount.

TABLE 6

altitude of lowest point C passed through by ray (h_C) (km.)	particle concentration at C relative to that at ground-level (n_C/n_G)	local scale height H (km.)	attenuation factor		
			λ 3500 ($A = 0{\cdot}589$/atm.)	λ 4000 ($A = 0{\cdot}345$/atm.)	λ 4500 ($A = 0{\cdot}214$/atm.)
0	1·0	8·3	0·00	0·00	0·00
10	$2{\cdot}7 \times 10^{-1}$	6·5	0·00	0·00	0·03
20	$7{\cdot}0 \times 10^{-2}$	6·4	0·08	0·23	0·40
30	$1{\cdot}5 \times 10^{-2}$	6·5	0·59	0·73	0·82
50	$8{\cdot}6 \times 10^{-4}$	10·0	0·96	0·98	0·99

Attention may be drawn here to the work of Störmer on the location of divided aurorae. The sunlit part of these aurorae (in which the principal excitation process is that under discussion) tend to end not exactly on the line of the earth's shadow but instead somewhat higher.* The results are rather scattered, but the average distance above does appear to be consistent with the result just reached.

3·3. *The observations*

Though, unfortunately, the absolute intensities of the bands appearing in the twilight flash have not yet been measured, the work of Dufay & Dufay (1947) at least enables crude limits (which for certain purposes are adequate) to be set.

Attention will be confined to three cases of interest for later applications. The methods used in estimating the limits will only be indicated, as they are not sufficiently precise to merit being given in detail.

In the following, h_L denotes the altitude of the lowest illuminated level along the line of observation and $E(h_L)$ the effective rate of photo/emission/cm.^{2} oblique column of the (0, 0) band λ 3914.

(i) For $h_L = 200$ km. the flash does not seem to have been recorded. Judging from Babcock & Jonson's spectrophotometric study of the light of the night sky (1941) this means that $E(200\,\text{km.})$ is always $< 10^7$/sec.

(ii) For $h_L = 60$ km. the flash has appeared prominently when conditions are disturbed, in spite of the strong background arising from Rayleigh scattering. In general, the intensity of this latter cannot be calculated accurately, as up to the present a complete theory has only been developed for a plane atmosphere (van de Hulst 1948). But the investigations of Hulburt (1938) suggest that for the particular h_L under consideration first order scattering is the most important;† and this can

* Cf. the diagram published by Störmer (1947).

† It should be noted that in estimating first-order scattering Hulburt adopted an atmosphere in which the particle concentration falls off too rapidly with altitude.

easily be computed. In order that the flash should show as it does it would appear that E(60 km.) is sometimes $> 10^9$/sec.

(iii) For $h_L = 100$ km. the flash is feeble or absent when conditions are quiet.* Scattering of the first order may here be negligible compared with that of higher orders. Some measure of the background intensity can, however, be obtained from the work of Scharonow (1944), who studied the degree of illumination at twilight sky, and of Megrelishvili (1946), who studied the colorimetry of the twilight sky. From this it was concluded that E(100 km.) is normally $< 10^8$/sec. (which is consistent with the fact that many of the true nocturnal emissions can be observed at the solar depression involved).

As need scarcely be emphasized the limits given are estimates only. Direct measurements would be invaluable and may lead to some revision of them. Nevertheless, as they are probably *conservative* it is thought that the main deductions to be made later will remain valid.

3·4. *Discussion*

A comparison between the observational estimates and the predictions of the two hypotheses under consideration will now be made. Throughout all the figures quoted will refer to the (0, 0) band λ 3914 and will be for a cm.^{2} oblique column.

It is immediately apparent that the Saha hypothesis in unsatisfactory. The greatest effective emission rate from process (1) that is given in table 1 is only $6{\cdot}3 \times 10^7$/sec.,† whereas it appears that E(60 km.) is sometimes greater than 10^9/sec. To yield such an emission it would be necessary to assume that on the occasion of a flash the ionization rate in the F_2 layer rises to be many times the normal. Further, the effective emission rate from process (2) is about $5{\cdot}7 \times 10^3 tn$/sec. (table 5, case 1 with $\beta = 20°$). This is more than the effective emission rate from process (1), if tn exceeds $\{6{\cdot}3 \times 10^7/5{\cdot}7 \times 10^3\}\{q_{mF_2}(\mathrm{N_2})/80\}$, that is, exceeds $1{\cdot}4 \times 10^2 q_{mF_2}(\mathrm{N_2})$—the quantity denoted by $q_{mF_2}(\mathrm{N_2})$ being here the rate of nitrogen ionization at the maximum of the layer during the period of the supposed enhancement of the incident flux of solar radiation.‡ It can easily be seen that tn must in fact be far greater than the comparatively small number given.§ Hence process (2) predominates. In view of this there is no justification for continuing to retain the artificial assumptions that had to be made to render the Saha hypothesis even plausible; when these are abandoned the contribution from process (1) is utterly negligible.

Finally, consider the Wulf-Deming hypothesis. So far it has only been shown that process (2) is much more rapid than process (1). It has still to be verified that even *it* can account for the flash. This presents no difficulty. $5{\cdot}7 \times 10^3 tn$/sec., the predicted effective emission rate, is so great that the required order of the intensity can be provided even if molecular nitrogen ions form only a small fraction of the total ion content of the upper atmosphere. Table 7 gives the values of $10^5 tn$ demanded by the limits set in § 3·3. Clearly the numbers are not unduly high. Their significance will be discussed later in connexion with the theory of the origin of the ionized layers.

---

* That quiet conditions are considered here, whereas in (ii) disturbed conditions are considered, is purely due to the particular information it is desired to extract.

† It will be recalled that this represents *a gross upper limit* for normal conditions.

‡ The corresponding average rate previously used is 80/cm.^{3}/sec.

§ If t is taken to be 200 km. the condition reduces to $n \gg 0{\cdot}7 q_{mF_2}(\mathrm{N_2})$, which is obviously satisfied.

TABLE 7

altitude range	number of nitrogen ions/cm.^{2} vertical column present at twilight ($10^5 tn$)
> 200 km. (which includes the *F* layer)	always $< 2 \times 10^8$
> 100 km. (which includes the *E* layer)	normally $< 2 \times 10^9$
> 60 km.	sometimes $> 2 \times 10^{10}$

4. THE THEORY OF THE FORMATION OF THE IONIZED LAYERS

4·1. *The F_1 and F_2 layers**

There now seems to be general agreement that the F_1 layer is produced by the ionization of atomic oxygen by solar radiation of photon energy 13·55 eV and above. It is perhaps worth summarizing briefly the arguments in favour of this. While these are not sufficient to prove the theory they are at least plausible. The rate of disappearance of the free electrons in the layer can be expressed as $\alpha n(e)^2$, where α is a constant (the effective recombination coefficient) and $n(e)$ is the electron density (cf. Appleton & Naismith 1935). In such circumstances Chapman (1931) has shown that the maximum of the layer occurs where the particle concentration $n_m(X)$ of the active atmospheric constituent X attains the value $\cos\chi/AH$, χ being the sun's zenith distance, A the absorption cross-section for the ionizing radiation, and H the local scale height. In the case of atomic oxygen it has been shown by quantal methods that at, and some distance above, the first ionization potential, A is about $4{\cdot}5 \times 10^{-18}$ cm.^{2} (Massey & Bates 1943); and reasons can be given for supposing that the approximations made in the course of the calculations do not introduce gross error as they would for many elements (Bates 1946*a*).† It can be seen that if the value of A quoted is adopted, if $\cos\chi$ is taken as 0·9, and if H the local scale height is assumed to be 30 km., then Chapman's formula yields $6{\cdot}7 \times 10^{10}$/cm.^{3} as the value for $n_m(X)$. This is reasonably consistent with the normally accepted total particle concentration in the F_1 layer, 1×10^{11}/cm.^{3} (Bates & Massey 1946). Further, the observed integrated rate of electron production in the layer is about $1{\cdot}8 \times 10^9$/cm.^{2} column/sec., which is not greatly different from 9×10^8/cm.^{2}/sec., the incident flux of photons in the relevant energy range from the sun, regarding it as a black body at 6000° K.‡

In the early stages of the development of the theory of the layers it was natural to attempt to attribute the F_2 layer to a process different from that responsible for the

---

* Attention will be confined throughout to electromagnetic radiation. There is, of course, evidence from eclipses indicating the incidence of some corpuscular radiation also, and it has been suggested that this may be important during magnetic storms (cf. Berkner 1939). The electromagnetic radiation, however, is the dominant ionizing agency in normal circumstances.

† Unfortunately, a limit cannot yet be assigned to the extent of the possible error. In view of the importance of a reliable knowledge of the cross-section the original calculation is being repeated with allowance for such effects as electron exchange.

‡ To make the two figures agree exactly it is only necessary to raise the black-body temperature by less than 200° K.

F_1 layer; and the most attractive possibility appeared to be the ionization of molecular nitrogen, the main gas in the atmosphere, beyond its lowest limit 15·6 eV. Little was known concerning the magnitude of the absorption cross-section, but crude estimates suggested it might be of the order required to provide a correctly located layer. The 6000° K black-body representation of the sun admittedly gives the flux of ionizing photons to be only $2{\cdot}7\times10^7$/cm.^{2}sec., considerably less than that needed for the observed integrated electron production rate $1{\cdot}4\times10^9$/cm.^{2} column/sec., but too much significance could not be attached to this. Mitra (1945) has pointed out that there is some reason for believing that the absorption cross-section at 15·6 eV is actually abnormally low. If this is the case ionization at the second limit 18·7 eV would have to be invoked. The absorption cross-section here is probably adequate, but the flux of suitable photons from the solar model used as a guide is only $7{\cdot}4\times10^4$/cm.^{2}/sec. A number of writers appear to advocate the hypothesis that it is molecules of nitrogen that are the parent particles from which the F_2 layer is formed.

More careful consideration of the problem shows that if certain upper atmospheric data are correct then the process proposed for the F_1 layer produces, *by itself*, sufficient ionization to give rise to the F_2 layer also, the presence of two layers being accounted for by the fact that the recombination coefficient changes with altitude.* On this view the F_1 layer is formed in the standard manner at the level of the maximum ionization rate, which, of course, occurs where the radiation is attenuated by a factor of e^{-1}; and the F_2 layer (for typical summer conditions) is formed at a considerable distance above where the radiation retains almost its incident intensity S_0, being reduced in fact by only the factor $\exp(-n_{mF_2}(O)/n_{mF_1}(O))$.† Hence $q_{mF_1}(O)$ and $q_{mF_2}(O)$ the electron production rates at the maxima of the two layers are respectively $S_0\exp(-1)\,n_{mF_1}(O)\,A$ and $S_0\exp(-n_{mF_2}(O)/n_{mF_1}(O))\,n_{mF_2}(O)\,A$. The ratio of $q_{mF_2}(O)$ to $q_{mF_1}(O)$ is

$$(n_{mF_2}(O)/n_{mF_1}(O))\exp\{1-(n_{mF_2}(O)/n_{mF_1}(O))\};$$

taking $n_{mF_1}(O)$‡ to be $6{\cdot}5\times10^{10}$/cm.^{3} and $n_{mF_2}(O)$ to be $1{\cdot}3\times10^{10}$/cm.^{3}, the value of this is 0·45. Now radio measurements show that $q_{mF_1}(O)$ is about 250/cm.^{3}/sec., so that on the above theory $q_{mF_2}(O)$ should be approximately 110/cm.^{3}/sec. Since the figure usually adopted for the electron production rate in the F_2 layer is only 80/cm.^{3}/sec., there is clearly no necessity to postulate an ionization process additional to that responsible for the F_1 layer; indeed, a slight adjustment of the data is required to prevent the predicted value of $q_{mF_2}(O)$ being rather greater than the value based on observation. Further, it was shown in § 3·4 that the twilight flash does not arise from process (1). The presumption that it did was one of the principal reasons for supposing that the solar emission in the 18·7 eV region might be greatly in excess of that from the black-body model; and if the emission is not in excess then the rate of ionization of molecular nitrogen is very small. Consistent with this is the minuteness of the concentration of N_2^+ ions present at twilight. Thus the integrated number/cm.^{2} column is at most 2×10^8 (see table 6), corresponding to perhaps only 10/cm.^{3} at the level of the maximum of the layer—a value many orders less than the total ion concentration. Indeed, it might be thought that here is conclusive proof that molecular nitrogen does not make any appreciable contribution to the ionization. For, from the formulae in § 3·1, it can readily be verified that even where the solar angle of depression is 15° the rate of formation of N_2^+ ions would be such that the almost complete absence of them that is demanded could not be achieved by any acceptable destructive process. However, this argument must be treated with a certain reserve, as it implicitly assumes that molecular nitrogen is still a considerable constituent of the atmosphere at altitudes of some 500 to 600 km. (which is where the ionization would occur). Now spectroscopic evidence leaves no doubt that on at least some occasions the nitrogen is partly in the atomic form.* The degree of the dissociation and the region where it occurs are unknown. Hence the possibility cannot be ignored that at the altitudes mentioned molecular nitrogen is rare.† If this is the case the formation of N_2^+ ions would virtually cease shortly after the sun sinks below the horizon, and by the time conditions are suitable for the observations on the negative band system to be made their concentration might be greatly diminished. Consider, for example, dissociative recombination

---

* This general idea appears to have been put forward first by Bradbury (1938) though on the somewhat dubious grounds that there could not be an absorption process so effective as to give rise to a layer at the low particle concentration involved. I am grateful to Dr D. F. Martyn for drawing my attention to Bradbury's work.

† For simplicity an isothermal atmosphere is used.

‡ To avoid inconsistencies in the subsequent treatment the value of n_{mF} (O) derived from the application of Chapman's formula is used here instead of the value normally assumed.

$$N_2^+ + e \rightarrow N' + N''. \tag{21}$$

It is conceivable that the rate coefficient associated with it is as high as 10^{-8}/cm.^{3}/sec.‡ Adopting this figure and taking the electron density to be 4×10^5/cm.^{3}, it can be seen that the mean lifetime of an N_2^+ ion is 250 sec.; and after say 3000 sec. the fraction of the original N_2^+ ions that would remain would be only 6×10^{-6}, which is as low as is required. Nevertheless, in spite of this it seems best, for the reasons given earlier, to assume that the ionization of molecular nitrogen is unimportant.

Before leaving the problem of the formation of the F_2 layer it is desirable to investigate the possible influence of atomic nitrogen. The basic information relevant for this purpose is as follows: the ionization potential is 14·46 eV; at the spectral head the absorption cross-section as determined by quantal calculations (Bates 1939, 1946), is $2{\cdot}2\times 10^{-17}$ cm.^{2}; the flux of ionizing photons available is $1{\cdot}8\times 10^8\ C(\nu)$/cm.^{2}/sec., where $C(\nu)$ is the correction factor to be applied to the standard solar model used (at 13·55 eV $C(\nu)$ is about 2, at 14·46 eV it is probably about the same or perhaps slightly greater).

Table 8 compares, for various degrees of dissociation D, the electron production rates in the F_2 layer due to atomic oxygen and to atomic nitrogen, the total particle concentration n_{mF_2} (which is also given) being adjusted so that the sum of these two rates remains at 80/cm.³/sec. In compiling the table diffusive separation of the atmospheric constituents was ignored, as was the small interaction of one absorption process on the other. Attention may be drawn to the fact that when the nitrogen is completely dissociated its concentration is just sufficient to attenuate the radiation ionizing, i.e. by the Chapman factor e^{-1}, at the layer maximum.*

---

* Cf. Dufay (1943*b*), Stebbing, Whitford & Swings (1945), Götz (1947), Bernard (1947), and Bates (1948, 1949).

† Auroral spectra, of course, show that there is some molecular nitrogen even at 1000 km., but its fractional concentration cannot be estimated reliably.

‡ It should be noted that the recombination coefficient observed by radio workers (which is much smaller than 10^{-8}/cm.^{3}/sec.) must not be identified with the recombination coefficient of a particular ionic constituent such as N_2^+ (cf. Bates & Massey 1947).

TABLE 8

fraction of nitrogen molecules dissociated (D)	electron production rates: atomic oxygen $q_{mF_2}(O)$ (per cm.³/sec.)	electron production rates: atomic nitrogen $q_{mF_2}(N)$ zero (per cm.³/sec.)	total particle concentration (n_{mF_2}) ($\times 10^{10}$/cm.³)
0·00	80	0	2·7
0·05	67	13	2·3
0·10	59	21	2·0
0·20	49	31	1·8
0·40	39	41	1·6
0·70	33	47	1·5
1·00	30	50	1·6

It is apparent from examination of table 8 that in no case is the value of n_{mF_2} required unacceptable—the lack of precision in the present knowledge of the upper atmosphere is such that the observed ionization in the F_2 layer can be reproduced irrespective of the form in which the nitrogen occurs. No final decision as to whether the parent particles responsible for the layer are solely oxygen atoms or are both oxygen atoms and nitrogen atoms is therefore possible; it will be noted, however, that the oxygen ionization is always a considerable fraction of the nitrogen ionization and indeed exceeds it unless D attains the value 0·4.

Despite the uncertainties several significant results emerge from the analysis that has been given.

(1) The ions formed in the layer are atomic and not molecular. Account must be taken of this in any study of the recombination processes operative (cf. Bates & Massey 1947).

(2) For low to medium D the layer is essentially of the low attenuation type, and it remains so partially even if D approaches unity. The characteristic of the extreme zero-attenuation layer is that q_m, the electron production rate at the maximum, is proportional to $n_m(X)$, the concentration of active particles, and does not vary specifically with χ, the solar zenith angle; this is to be contrasted with the normal attenuation layer in which q_m is in general† a function of both $n_m(X)$ and χ. The recognition from the radio records of the distinctive behaviour to be expected is, of course, rendered difficult by a number of complicating effects among which may be

---

* The scale height here was taken to be 30 km. This rather low value was adopted because of the comparatively rapid decrease in gas density between the two layers that is required to give the observed ionization rates. The general results exhibited by the table are not sensitive to the scale height assumed.

† It is only in special circumstances (constant recombination coefficient; disturbing factors absent) that $n_m(X)$ and χ are simply related so that q_m can be expressed in terms of either.

mentioned the remarkable tidal motions recently discovered by Martyn (1947); Mohler (1940), however, claims that there is some evidence favouring it.

(3) The same (or very nearly the same) region of the solar spectrum is responsible for the production of the F_1 and the F_2 layers. Any variation of the emission will affect both layers. Kiepenheuer (1945) has published a number of diagrams that illustrate beautifully the expected correlation.* The behaviour of the relatively simple and well-understood F_1 layer should therefore provide some information regarding one of the many factors which govern the F_2 layer.

It might be supposed at first that the theory is in conflict with the variation of the two layers throughout the sunspot cycle; for as is well known (cf. Berkner 1939) the electron density in the F_1 layer increases by a factor of about 1·6 from sunspot minimum to sunspot maximum, whereas the corresponding factor for the F_2 layer is about 4. However, no contradiction is necessarily involved. Thus in the F_1 layer the electron loss rate depends on the second power of the electron density; but in the F_2 layer it probably depends on the first power (cf. Bates & Massey 1947) which, of course, would lead to a greater sensitivity to the ionization rate. Again, Bates & Massey (1947) have proposed tentatively that the recombination mechanism operative is the following:

$$O^+ + XY \rightarrow O + XY^+, \quad XY^+ + e \rightarrow X' + Y'. \tag{22}$$

On this theory the electron loss rate in the F_2 layer is proportional to the concentration there of the unidentified molecule XY which might conceivably be less at sunspot maximum than at sunspot minimum. And there are other obvious possibilities which need not be enumerated.

(4) Finally, it is worth making some mention of the electron production rates in the two layers. q_{mF_1} was taken as 250/cm.^{3}/sec. and q_{mF_2} as 80/cm.^{3}/sec. These representative values were based respectively on electron densities of $2{\cdot}5 \times 10^5$/cm.^{3} and 1×10^6/cm.^{3}, and on recombination coefficients of 4×10^{-9} cm.^{3}/sec. and 8×10^{-11} cm.^{3}/sec. The electron densities are not in doubt. The recombination coefficient in the F_1 layer is also probably reliable; the figure quoted is that adopted by Bates & Massey (1946) after consideration of the radio evidence available; it lies satisfactorily among more recent determinations.† But as regards the recombination coefficient in the F_2 layer there is still controversy. The value 8×10^{-11} cm.^{3}/sec. was derived by Appleton (1937) from the asymmetry in the diurnal ionization variation; Martyn (1947), however, suggests that tidal motion is so large that this procedure is unjustifiable, and Allen (1946) and Woolley (1946) favour the very small value 1×10^{-11} cm.^{3}/sec.‡ The present investigation shows that the electron

---

* As need scarcely be mentioned this agreement does not *prove* the theory.

† Rydbeck (1946) gives 6×10^{-9} cm.^{3}/sec., Alpert & Einberg (1947) give 7×10^{-9} cm.^{3}/sec., and Denisse, Seligmann & Gallet (1947) give 2×10^{-9} cm.^{3}/sec. All were obtained from eclipse measurements.

‡ According to Martyn the reason for the difference between this and the nighttime value of 3×10^{-10} cm.^{3}/sec., deduced directly from the decay rate, is that at sunset the ions and electrons move downwards under tidal influence to a region where recombination is comparatively rapid.

production rate of 80/cm.^{3}/sec. associated with Appleton's coefficient is reasonable; a much lower rate would necessitate the particle concentration in the layer being considerably less than $2{\cdot}7 \times 10^{10}$/cm.^{3}.

4·2. *The E layer*

The absorption cross-section of atomic oxygen seems to be such that photons of energy between 13·55 eV and well over 25 eV cannot penetrate much deeper than the F_1 layer (Bates & Massey 1946). As a consequence of this there is a severe limitation in the processes available for E-layer ionization. Molecular oxygen is the only major atmospheric constituent with ionization potential less than the lower limit; ionization of it, which requires only 12·2 eV energy, is therefore a natural process to consider. Above the upper limit the photon flux from the 6000° K black-body model is very many orders of magnitude smaller than the observed ionization rate 6×10^8/cm.^{2} column/sec. This might at first lead one to suppose that the spectral region concerned could be ignored. However, recent work by Bondi, Hoyle & Lyttleton (1947) has shown that the sun's emission of photons of energy a few hundred electron volts may actually be of the required intensity.

Several groups (cf. Wulf & Deming 1938; Bates & Massey 1946) have given consideration to the molecular oxygen theory. As has been realized for some time this is not attractive in its simplest form. The absorption cross-section at the first ionization potential 12·2 eV seems to be very low, perhaps only 10^{-20} cm.* As a consequence the location requirement does not appear to be met. The difficulty is particularly acute when account is taken of the allotropic form of the oxygen. This has been investigated by a number of workers. The treatment of Wulf & Deming (1938) is the most satisfactory, and though not precise gives the main trend. When adjustment is made to the results originally given, to allow for the fact that the total particle concentration in the basic atmosphere adopted falls off too rapidly with altitude, it appears that dissociation commences sharply at about 105 km. and is almost complete at 120 km.

Though the absorption spectrum of molecular oxygen shows only a feeble continuum in the energy range 12·2 to 13·55 eV, it shows superimposed a number of strong bands in which the absorption cross-section is some 4×10^{-17} cm.^{2} (Schneider 1940). Nicolet (1945) has made the interesting suggestion that these may lead to pre-ionization and that this may be the mechanism responsible for the formation of the E layer. If the scale height associated with the oxygen molecules is denoted by $H(\mathrm{O_2})$, then $n_{mE}(\mathrm{O_2})$, their concentration at the layer maximum, is as usual $\cos\chi/AH(\mathrm{O_2})$. $\cos\chi$ can be taken to be 0·9 and A has just been given. Instead of adopting any particular value for $H(\mathrm{O_2})$ it is best to assume that $n_{mE}(\mathrm{O_2})$ can be equated also to $n_D(\mathrm{O_2}) \exp(-z/H(\mathrm{O_2}))$, where $n_D(\mathrm{O_2})$ is the concentration of oxygen molecules where dissociation first becomes appreciable and z is the distance this is below the layer maximum. From the two expressions $H(\mathrm{O_2})$ can be determined and the figure obtained can then be used in the evaluation of $n_{mE}(\mathrm{O_2})$. Now $n_D(\mathrm{O_2})$ is some $9 \times 10^{12}/\text{cm.}^3$ and z some 15 km. Hence it can be seen that $H(O_2)$ and $n_{mE}(O_2)$ are about 3·2 km. and $8 \times 10^{10}/\text{cm.}^3$ respectively.* Great significance cannot be attached to these values in view of the approximation used in deriving them, but at least they appear reasonably consistent with the calculations of Wulf & Deming (modified as described above). It might be thought that the smallness of $H(O_2)$† is in conflict with Appleton's measurements (1937) which give the scale height associated with the profile of the under-surface of the layer just below the maximum to be 11·4 km. But this is not necessarily the case, since for the absorption process under consideration the cross-section varies rapidly with frequency thereby causing a spread in the ionization. Other problems regarding the scale height will be taken up later.

---

* The second ionization potential, 16·1 eV, is of course above the critical energy 13·55 eV. Confirmatory measurements on the absorption cross-section are needed. The value quoted is an estimate only (cf. Mitra, Bhar & Ghosh 1938, where a somewhat lower value is suggested).

As the incident flux of photons in the 12·2 to 13·55 eV energy range is estimated from the standard solar model to be some $9 \times 10^9/\text{cm.}^2/\text{sec.}$, and as the absorption bands are quite wide (Schneider 1940), the electron production rate from pre-ionization may well have the required value $6 \times 10^8/\text{cm.}^2$ column/sec. Nicolet (1945) pointed out that the lines L_γ (λ 973) and L_ξ (λ 931) of hydrogen coincide with two of the bands.

The high-energy photon theory has been investigated by Hoyle & Bates (1948). It was shown that if the sun emitted photon groups at about 425 eV or at about 1300 eV an ionized layer would be produced at the required altitude. When consideration was given to the ionizing effect of the secondary electrons ejected in either case it was found that the 1300 eV photons were not altogether satisfactory. The 325 eV photons did not, however, appear to lead to any discrepancies. Further, recent developments in solar physics suggests that their flux from the corona may be sufficient to yield the observed degree of ionization (cf. Hoyle 1949).

Though the pre-ionization theory and the high-energy photon theory (which in future will be referred to as theories A and B respectively) are plausible, neither can as yet be regarded as other than speculative since both depend on an unproved assumption. Thus regarding theory A, there is no direct evidence that molecules in the excited state resulting from band absorption do in fact ionize spontaneously instead of becoming stabilized by photon emission; it can only be said that, as far as can be judged, the relevant selection rules permit this to occur. And regarding theory B, lack of knowledge of the chemical composition of the corona makes it impossible to state firmly that the flux of the 325 eV photon group is as great as required; all that is known is that such a flux is a possibility. But as no acceptable alternative has yet been put forward in spite of the efforts of many workers it is reasonable to assume that one or other (or a combination of both) is correct. Doubtless the matter will be decided finally by laboratory investigations on pre-ionization in oxygen, by further advances in information about the corona, and by using rocket projectiles to study directly the radiation responsible for the E layer. Meanwhile, it is worth mentioning briefly a few less direct methods of approach. This can most conveniently be done under a number of heads.

---

* For comparison it may be noted that at 120 km. it is generally taken that the scale height of the main atmosphere is somewhat above 10 km. (as judged from the temperature), and that the total particle concentration is about $6 \times 10^{12}/\text{cm.}^3$.

† A much larger value than that given here is deduced by Nicolet (1945) by a different argument. His treatment involves apparently independent assumptions regarding the absorption cross-section for the photo-dissociation of molecular oxygen and regarding the altitude of the level at which dissociation is a maximum. In fact, the two assumptions are inconsistent.

Eclipses

The photons invoked in theory A originate from closer to the sun's surface than those invoked in theory B. It may, therefore, be possible to distinguish between the two theories by careful examination of the change in the layer during the course of an eclipse. No systematic study of the records has yet been made. Rydbeck (1946), however, claims that E-layer measurements at the time of the Swedish eclipse of July 1945 gave some indication of ionizing radiation from the corona. This favours theory B. The position is complicated by the fact that similar indications were claimed for the F_1 layer. Clearly further work is required.

Sunspot cycle

The E and the F_1 layers change by almost equal factors from sunspot minimum to sunspot maximum (cf. Appleton & Naismith 1939); they also show much the same dependence on sunspot number (cf. Waldmeier 1944). Now there can be little doubt that the F_1 layer is produced by photons of energy 13·55 eV and somewhat greater. The photons of theory A (12·2 to 13·55 eV) lie on the opposite side of the vital Lyman limit to these, and the photons of theory B (325 eV) are far more energetic. This makes it difficult to decide which theory the resemblance between the layers favours; but solar physicists may eventually provide an answer.*

Altitude distribution of ionization

Probably fair representation of the ionization distribution below the maximum of the layer can be obtained from either theory. On theory A, as already mentioned, the distribution is governed by the extent of the variation of the absorption cross-section. Insufficient information is available to allow a reliable prediction to be made, but it is possible that the theory can yield not only the distribution just below the layer maximum (where the associated scale height is 11·4 km.), but that it can also yield the distribution well below the layer maximum (where the associated scale height is thought to be less than 11·4 km.). On theory B the distribution depends to some extent on the degree of non-homogeneity of the high-energy photon group. It does not, however, differ markedly from the standard Chapman distribution with associated scale height that of the main atmosphere (which is about 10 km.). For example, in a representative case Hoyle & Bates found that the electron density fell by factors e^{-1} and e^{-2} in distances of 17 and 22 km. respectively below the maximum; the corresponding distances in a standard Chapman distribution are 15 and 19 km. Theory A seems on the whole to be more satisfactory than theory B—but, of course, its apparent greater flexibility is solely due to lack of basic data.

Above the layer maximum the predictions of the two theories differ markedly. Owing to the very rapid fall off with altitude of the molecular oxygen concentration theory A yields very little ionization in the region between the E and F_1 layers. In

* The author actually consulted two solar physicists; their views were diametrically opposed.

contrast, on theory B there would be considerable ionization. Thus if at 170 km. (that is the midway level) the total particle concentration and the effective recombination coefficient are taken to be 6×10^{11}/cm.^{3} and 7×10^{-9} cm.^{3}/sec. respectively,* then the electron density would be smaller than that at maximum of the E layer by only a factor of about 0·6. At present there is no information on the actual value of the electron density. It is possible that information on it could be obtained by the study of the absorption of radio waves.

Variation with solar zenith angle

Chapman (1931) has shown that the altitude of the maximum of an ionized layer is related to the solar zenith angle by the equation

$$z_m = C - H(i) \log(\cos\chi), \tag{23}$$

C being a constant and $H(i)$ the scale height of the atmospheric constituent responsible for the ionization. According to whether theory A or B is adopted $H(i)$ is the scale height of molecular oxygen in the transition region or is the scale height of the main atmosphere. In the first case its magnitude is probably only about 3·2 km. (*vide supra*), but in the second case its magnitude is 10 km. or rather more. It would be expected therefore that a guide as to which theory is correct would be provided by the variation of z_m with χ (which, of course, is contained in measurements made at different seasons, at different hours of the day, or at different latitudes). Unfortunately, in spite of the considerable difference in the two scale heights, discrimination is not easy.

Consider the seasonal effect. Appleton & Naismith (1940), working at Slough, obtained the results given in table 9. Substitution in (23) immediately yields

$$1{\cdot}2_3 H(i) = 14\,\text{km.} \quad \text{or} \quad H(i) = 11{\cdot}4\,\text{km.}$$

TABLE 9

season	χ (°)	*zm* (km.)
summer	28	120
winter	75	134

This is in excellent agreement with theory B, and at first seems to be in serious conflict with theory A. However, allowance must be made for the possible seasonal change in the O-O$_2$ distribution. Though an exact investigation of this is scarcely feasible at present useful preliminary results can readily be obtained by a crude approximate calculation. In this attention will be confined to altitudes close to 134 km. Let x denote the winter to summer ratio of the molecular oxygen concentration at some particular level, r the corresponding ratio of the dissociation rates/unit concentration and d that of the durations of day light. Clearly,

$$xrd \simeq 1. \tag{24}$$

Owing to the dependence of the absorption cross-section on frequency the dissociation rate/unit concentration cannot be expressed in exact analytical form; but over a limited altitude range it can be written as approximately proportional to some inverse power p of the amount of molecular oxygen traversed by the radiation before reaching the level concerned. Hence from (24),

---

* The figures are estimates based on the geometric means of the corresponding figures in the two layers.

$$x\{x(\sec\chi_W/\sec\chi_S)\}^{-p}\,d \doteq 1, \tag{25}$$

the subscripts to χ indicating the season. At the latitude of Slough (51·5° N) d is $0{\cdot}5_5$, $\sec\chi_W$ is 3·86 and $\sec\chi_S$ is 1·13. In connexion with their study of the photochemistry of atmospheric oxygen Wulf & Deming (1936) have published curves illustrating how the rate of absorption of the solar radiation varies with altitude. From these it can be deduced that p is about $0{\cdot}4_5$. Hence it can be seen from (25) that x is of order 8. This corresponds to a level of given molecular oxygen concentration rising a distance $2{\cdot}1H(O_2)$ from summer to winter. The change in the altitude of the E layer itself should therefore be about $3{\cdot}3H(O_2)$ km., that is (taking $H(O_2)$ to be 3·2 km. as before) about 11 km. In view of the many simplifications introduced this is in not unreasonable agreement with the observed 14 km.

Little need be said regarding the diurnal and latitudinal variations. The smallness of the altitude changes in the former (cf. Hulburt 1939) would, however, seem to favour theory B. As far as the latter is concerned measurements at various stations yield altitudes ranging from as low as 100 km. to as high as 140 km. (cf. Hulburt 1939), which would suggest a great scale height; but this must be treated with reserve, since the published results are not always strictly comparable due to different criterion being adopted; and of course the O-O_2 distribution depends on the latitude. An investigation on both variations specifically designed to obtain information regarding the scale heights would be valuable.

Spectroscopic observations

Finally, the spectroscopic observations have to be considered. Theories A and B have a very distinctive difference; on the former only O_2^+ ions are produced, but on the latter O^+ ions and (more important) N_2^+ ions are also produced at rates which may be presumed to be approximately proportional to the fractional concentrations of the parent neutral particles. If the number of N_2^+ ions/cm.^{3} at the maximum of the layer during twilight is denoted by $n_{mE(t)}(N_2^+)$, then for a Chapman distribution* the total number/cm.^{2} column is given by $\{2\pi\exp(1)\}^{\frac{1}{2}}Hn_{mE(t)}(N_2^+)$ which reduces to $4{\cdot}1Hn_{mE(t)}(N_2^+)$. Now in § 3·4 it was shown that this same number is normally less than 2×10^9. Hence

$$n_{mE(t)}(N_2^+) < 5\times10^8/H/\text{cm.}^3 < 5\times10^2/\text{cm.}^3,$$

with H as usual taken to be 10 km. The total ion concentration at the maximum of the layer just after sunset is about 2×10^4/cm.^{3}—some 40 times greater than the limit just deduced. Theory A is obviously in no difficulty. There is some doubt, however, as to whether the comparative rarity of the N_2^+ ions can be reconciled with theory B. It is clearly important to investigate this. Assuming the theory to be valid the rate of production of N_2^+ ions along a line of observation elevated at an angle β above the horizon can be computed for any given solar depression α by using the formulae in § 3·1. If in the region concerned the nitrogen remains mainly molecular as would seem reasonable from the investigations of Harang (1945) then with $\alpha = 12\frac{1}{2}°$ and $\beta = 20°$ (which are sufficiently representative values) it can be shown that the production rate attains a maximum of order 5/cm.³/sec. at an altitude of some 200 km. Because of the lower electron density, dissociative recombination (21) is relatively less important than in the *F* layer and is probably insufficient to maintain the equilibrium number of N_2^+ ions at 5×10^2/cm.³. The most promising process is that of charge transfer

---

* The actual distribution is unknown, but the error introduced by assuming it to be Chapman-like is not serious.

$$N_2^+ + O \rightarrow N_2 + O^+. \tag{26}$$

Taking the oxygen atom concentration to be about 5×10^{10}/cm.³, it can be seen that for the N_2^+ ions to be removed rapidly enough the associated coefficient γ must be at least 2×10^{-13} cm.³/sec. The somewhat meagre information available from theoretical and experimental studies of charge transfer collisions do not exclude this value. However, account must be taken of the fact that the flash is sometimes emitted strongly from the neighbourhood of the 100 km. level. On such occasions the N_2^+ concentration must increase to 5×10^3/cm.³ or more (cf. § 3·4 for basic data). It is at least plausible to assume that incident charged particles are responsible for the additional ionization (cf. § 2). These must be a weak source compared with the normal solar electromagnetic radiation, as otherwise the remarkable regularity of the *E* layer could not be preserved, and, furthermore, the negative band system would appear prominently in the night-sky light;* it is difficult to fix an upper limit to the total rate of ion production permissible, but perhaps 30/cm.³/sec. is a reasonable provisional figure to adopt,† with the same assumptions as in theory B the number of N_2^+ ions formed is therefore 20/cm.³/sec. or less. Now the N_2^+ ions are constantly being lost through charge transfer collisions with oxygen atoms. Since the concentration of these latter at the level concerned is about 2×10^{12}/cm.³, it can be seen that unless the coefficient is lower than 2×10^{-15} cm.³/sec. the required concentration of N_2^+ ions cannot be established.‡ In view of this, the demand of theory B that γ be greater than 2×10^{-13}/cm.³/sec. is clearly unacceptable.

While the simple argument that has been given certainly favours the exclusion of theory B it should not be regarded as conclusive. By making certain hypotheses regarding the variation of the composition of the atmosphere with altitude and introducing in the upper regions alternative loss processes such as

$$N_2^+ + N \rightarrow N_2 + N^+, \tag{27}$$

with assumed high rate coefficients it is possible to remove the discrepancy. Proposals so conjectural do not, however, justify lengthy development at present. An investigation on the effectiveness of charge transfer collisions should assist in clarifying the issue. It would also be of value to have proper information on the ratio of the intensities of the flash when strong and when feeble or apparently absent. The value of this ratio has throughout been taken as 10, which is thought to be an underestimate—if it should prove to be a gross underestimate theory B would become difficult to sustain.

---

* *Vide infra.*

† For comparison it may be noted that at midday q_{mE} is about 220/cm.³/sec.

‡ As can readily be verified if γ is as great as 2×10^{-15} cm.³/sec. and N_2^+ ions formed near the maximum of the layer during the day would rapidly disappear after sunset. This is why attention is confined to the high levels that continue to be reached by ionizing radiation. That γ is of such a magnitude is, however, uncertain (cf. § 5).

To summarize, it may be said that the various aspects of the layers that have been discussed do not, unfortunately, enable any final decision to be reached at present as to whether it is theory A or theory B that is correct; but the former seems to be rather more favoured than the latter. When further evidence in the directions that have been indicated becomes available it may be possible to form some definite conclusion—though it must be borne in mind that overlooked subordinate effects can readily cause a misinterpretation of the detailed properties of any upper atmospheric phenomena.

5. THE NIGHT-SKY LIGHT

It has been suggested by Mitra (1943) and Ghosh (1943) that the emission of the nitrogen-band systems from the upper atmosphere during the night arises from recombination of the charged particles in the ionosphere. This view has recently been criticized by Bates (1948), but the whole subject is so complex and imperfectly understood that it is perhaps worth drawing attention to the fact that the present investigation provides more conclusive evidence than any previously available that some other source must be sought. Thus it has been shown in § 3·4 that just after sunset the total number of N_2^+ ions/cm.^{2}/column is normally not more than 2×10^9. Even neglecting loss by process (21), (26) and (27) (which cannot yield band emission), it is apparent that this number is totally inadequate to account for the observed intensity. Consider, for example, the Vegard-Kaplan system alone, and suppose that in the course of a night one-half of the ions contribute to the thirty individual bands of it that have been reported by Cabannes & Dufay (1946). It can readily be seen that the maximum average rate of photon emission in each could only be about 10^3/cm.^{2} column/sec. which, of course, is far below what is required.

To account for the occasional great prominence of the twilight flash it was suggested in § 2·1 that additional N_2^+ ions might be produced by incident charged particles. Now these presumably might enter the earth's atmosphere during darkness. It is of interest therefore to investigate whether they could make an appreciable contribution to the true nocturnal emission. The ionization rate used in § 4·2* was 30/cm.^{3}/sec., which corresponds to about 10^8/cm.^{2} column/sec. The photon intensity resulting from excitations would be expected to be comparable and would thus be observable—indeed, it should be possible to detect the emission associated with a considerably smaller ionization rate. Among the radiations that would arise would be bands of the negative system of nitrogen. As mentioned in § 1 there is some evidence for the sporadic appearance of these. It would be a useful check if a correlation could be established between their intensity in the twilight flash and their intensity in the night-sky spectrum proper. Data on the maximum absolute magnitude of their intensity in this latter would also be valuable, as it would give a guide to the maximum ionization rate which in turn would enable a proper upper limit to be attached to γ, the coefficient associated with process (26). If this should prove to be much smaller than the 2×10^{-15} cm.^{3}/sec. adopted in § 4·2 (as might well happen) then even at *low* levels the day produced N_2^+ ions would disappear so slowly that theory B would encounter difficulties additional to those already mentioned. These could not be avoided by the elaborations that were proposed in connexion with the problem of the disappearance of the N_2^+ ions formed during twilight at *high* levels.

---

* It will be recalled that this ionization rate was chosen solely to satisfy theory B as far as possible; the actual ionization rate may be less by an order of magnitude.

6. SUNLIT AURORAE

Sunlit aurorae are basically very similar to the twilight flash. Their remarkable brilliance is due to the incoming corpuscles *sensitizing* the atmosphere by producing along their track N_2^+ ions which, under the influence of the solar radiation, become powerful sources of the negative bands. The high degree of the sensitization that must occur can readily be demonstrated. Consider first the photon emission I_C that results from the simultaneous ionization and excitation of nitrogen molecules. Once equilibrium is established the total rate of formation of N_2^+ ions must equal their recombination rate.* Now only a fraction of the N_2^+ ions would initially be in the $B\,^2\Sigma_u^+$ state. Hence I_C must be less than the recombination rate; that is

$$I_C < \alpha n(e)\, n(\mathrm{N}_2^+)/\text{cm.}^3/\text{sec.}, \tag{28}$$

where α is the recombination coefficient between electrons and N_2^+ ions, and the n's represent their concentrations. For resonance excitation the photon emission I_R is given simply by

$$I_R = Pn(\mathrm{N}_2^+)/\text{cm.}^3/\text{sec.}, \tag{29}$$

P being defined in § 3·2. Thus

$$I_R/I_C > \{P/\alpha n(e)\}. \tag{30}$$

The value of P has been shown to be about 0·13/sec.; that of α can scarcely be greater than 10^{-8}/cm.^{3}/sec. It follows that

$$I_R/I_C > \{10^7/n(e)\}. \tag{31}$$

Unless $n(e)$ is more than 10^7/cm.^{3} (which is a very high concentration†), I_R is therefore greater than I_C; and for normal sunlit aurorae it is greater by many orders. The divided aurorae reported by Störmer (1947) well illustrate this difference between I_R and I_C. Above the shadow line there is a section that is rendered luminous by resonance excitation in the sensitized part of the atmosphere through which the auroral stream passes. Immediately below the shadow line is a dark section arising because the relative inefficiency of excitation by collisions is such that at the small gas densities involved no detectable photon emission results. Finally, if the auroral stream penetrates to levels where the gas density is sufficiently high a third section, luminous like the first, naturally appears.

In conclusion, I would like to thank Professor H. S. W. Massey, F.R.S., for his continued interest and for helpful discussions.

---

* At great altitudes loss of N_2^+ ions by charge transfer is negligible.

† It should be noted too that it is a lower limit.

7. APPENDIX

The spontaneous transition probability associated with emission of the negative bands of nitrogen

While the calculation of spontaneous transition probabilities associated with electronic transitions in molecules is of necessity based on simplifying assumptions, the pioneer work of Mulliken (1939*a*) indicates that the results obtainable are usually not greatly in error. The fundamental quantal formulae involved are, of course, the same as in the corresponding atomic problem. Regarding the nuclei as fixed the total *f*-value for the transition (which it is convenient to evaluate first) is given by†

$$f = \{1/3Ra_0^2\lambda\}D, \tag{A. 1}$$

where R is Rydberg's constant, a_0 is the radius of the first Bohr orbit, λ is the wavelength of the radiation absorbed and D is the dipole moment factor,

$$\left|\int_\tau X_i(\mathbf{r})\,\Sigma\mathbf{r}X_f^*(\mathbf{r})\,d\tau\right|^2, \tag{A. 2}$$

X_i and X_f being respectively the initial and final normalized electronic wave functions and $\mathbf{r}$ the electronic position vector. Numerical substitution yields

$$f = \frac{304}{\lambda}D, \tag{A. 3}$$

with λ now in Ångström units and D in atomic units. A, the spontaneous transition probability, is related to the *f*-value by†

$$A = \frac{8\pi^2e^2}{mc\lambda^2}f\,\text{sec.}^{-1}, \tag{A. 4}$$

m, c and e having their standard significance. This at once gives

$$A = \frac{6\cdot6\times10^{15}f}{\lambda^2}\,\text{sec.}^{-1}, \tag{A. 5}$$

λ as in (3) being in Ångström units.

For the transition under discussion the ground state $X\,^2\Sigma_g^+$ has the configuration $KK(\sigma_g 2s)^2(\sigma_u 2s)^2(\pi_u 2p)^4(\sigma_g 2p)$, and the excited state $B\,^2\Sigma_u^+$ has (at moderate internuclear separation) the configuration $KK(\sigma_g 2s)^2(\sigma_u 2s)(\pi_u 2p)^4(\sigma_g 2p)^2$ (cf. Mulliken 1932). Neglecting the molecular orbitals not taking any active part, the transition may be regarded simply as

$$(\sigma_u 2s)^2(\sigma_g 2p)-(\sigma_u 2s)(\sigma_g 2p)^2. \tag{A. 6}$$

By writing the wave functions in their determinental form it can be seen that this can be further reduced to

$$(\sigma_u 2s)-(\sigma_g 2p). \tag{A. 7}$$

---

† The orbital degeneracy factors are omitted as in the transition considered here they equal unity.

The wave functions of the $(\sigma_u 2s)$ and $(\sigma_g 2p)$ molecular orbitals can be approximately represented by suitable linear combinations of atomic wave functions. Following the usual procedure we have

$$X(\sigma_u 2s \mid \mathbf{r}) = \{\phi(2,0,0,A \mid \mathbf{r}) - \phi(2,0,0,B \mid \mathbf{r})\}/2^{\frac{1}{2}}(1 - S(2,0,0))^{\frac{1}{2}}, \qquad (\text{A. }8)$$

$$X(\sigma_g 2p \mid \mathbf{r}) = \{\phi(2,1,0,A \mid \mathbf{r}) - \phi(2,1,0,B \mid \mathbf{r})\}/2^{\frac{1}{2}}(1 - S(2,1,0))^{\frac{1}{2}}. \qquad (\text{A. }9)$$

Here the ϕ's are the normalized atomic wave functions with principal, azimuthal and magnetic quantum numbers n, l and m respectively as indicated, and with the nucleus to which they are referred, denoted by the letters A and B. The S's which are introduced in the normalization are the overlap integrals given by

$$S(n,l,m) = \int_\tau \phi(n,l,m,A \mid \mathbf{r})\,\phi(n,l,m,B \mid \mathbf{r})\,d\tau. \qquad (\text{A. }10)$$

Substitution of (A. 8) for X_i and (A. 9) for X_f in (A. 2) gives

$$D = \{P(2,0,0;\,2,1,0) - T(2,0,0;\,2,1,0)\}^2/\{1 - S(2,0,0)\}\{1 - S(2,1,0)\}, \qquad (\text{A. }11)$$

where
$$P(2,0,0;\,2,1,0) = \int_\tau \phi(2,0,0,A \mid \mathbf{r})\,\mathbf{r}\phi(2,1,0,A \mid \mathbf{r})\,d\tau$$
$$= \int_\tau \phi(2,0,0,B \mid \mathbf{r})\,\mathbf{r}\phi(2,1,0,B \mid \mathbf{r})\,d\tau \qquad (\text{A. }12)$$

and
$$T(2,0,0;\,2,1,0) = \int_\tau \phi(2,0,0A \mid \mathbf{r})\,\mathbf{r}\phi(2,1,0,B \mid \mathbf{r})\,d\tau$$
$$= \int_\tau \phi(2,0,0,B \mid \mathbf{r})\,\mathbf{r}\phi(2,1,0,A \mid \mathbf{r})\,d\tau. \qquad (\text{A. }13)$$

To complete the numerical work it is necessary to choose some form for the atomic wave functions. Two different approximations were taken:

I. Slater type functions with effective nuclear charge 4·08e.

II. Hartree-Fock type functions (as calculated by Brown, Bartlett & Dunn 1933) fitted by exponentials.

The first of these approximations is probably too compact and the second too diffuse.

Using the ϕ's as described the evaluation of P can be carried out immediately. The evaluation of the S's and of T requires the introduction of elliptic co-ordinates and leads to rather cumbersome expressions which need not be given here; the internuclear distance appears explicitly and was taken to be $1{\cdot}117 \times 10^{-8}$ cm., the value for the *ground* state (from Herzberg 1939). Table A. 1 gives the magnitude of

Table A. 1

	approximation I (atomic units)	approximation II (atomic units)
$S(2,0,0)$	$+0{\cdot}41$	$+0{\cdot}44$
$S(2,1,0)$	$-0{\cdot}33$	$-0{\cdot}23$
$P(2,0,0;\,2,1,0)$	$+0{\cdot}71$	$+0{\cdot}82$
$T(2,0,0;\,2,1,0)$	$+0{\cdot}21$	$+0{\cdot}27$

the various integrals as finally computed. From these the dipole moment factors D and the f-values can be derived using (A. 11) and (A. 3). They are given in table A. 2 with, in addition (under approximation III), the results obtained if it is assumed that the configuration of $N_2^+(B\,^2\Sigma_u^+)$ is $KK(\sigma_g 2s)^2(\sigma_u 2s)^2(\pi_u 2p)^4(\sigma_u 2p)$ (which is probably true at large internuclear separation (cf. Mulliken 1932)) so that the transition is essentially

$$(\sigma_g 2p) - (\sigma_u 2p), \tag{A. 14}$$

and thus is of the resonance type for which, as Mulliken (1939*b*) has pointed out, D is given by the simple expression

$$(r_e/2)^2/\{1 - S(2, 1, 0)^2\}. \tag{A. 15}$$

Approximation III should be given less weight than approximations I and II.

TABLE A. 2

	approximation I	approximation II	approximation III
D (atomic units)	0·32	0·44	1·22
f-value*	0·025	0·034	0·095

* In calculating the f-values λ was taken to be 3914 Å (*vide infra*).

No experimental determination of the f-value associated with the negative bands has been made, but White (1940) has carried out quantitative work on the corresponding system of the cyanide radicle (which is isoelectronic with the nitrogen molecular ion). Briefly, his experiment consisted of measuring the amount of band absorption produced by the cyanide radicles formed through the complete dissociation, by means of an electrical discharge, of the cyanogen molecules in a cyanogen-argon mixture. Careful consideration was given to the timing. The f-value obtained was 0·026. White stressed that this should be treated as a lower limit, since it was possible that some of the cyanide radicles themselves might be dissociated. He suggested that the correct figure might be several times greater, perhaps about 0·1. In a further experiment he investigated the thermal equilibrium of $(CN)_2 \rightleftharpoons 2CN$ by determining the CN concentration from the band absorption. He was able to derive in this way the dissociation energy of $(CN)_2$. The figure reached naturally depends on the f-value assumed. Because of the uncertainty in his measurements of this quantity he was only able to show that the dissociation energy must lie between 138 and 152 kcal. The correct figure is still uncertain, and White's result was the subject of some criticism at the time by Herzberg (1942) and by Robertson & Pease (1942). However, it is satisfactory to note that 140 kcal. (which is within the range given) is favoured by Gaydon (1947) in his recent critical survey of the position. It may be remarked that the f-value which, with the equilibrium data of White, yields this dissociation energy is 0·035. Too much significance should not, however, be attached to this. Some further evidence on the CN transition is provided by the work of Roach (1939) on the abundance of molecules in the solar atmosphere. While not sufficient to enable the f-value to be estimated this at least supports the view that it is not abnormally low.

Returning to N_2^+, the data given above suggests that the f-value for the negative bands is about 0·04 (within perhaps a factor of three either way). For the purpose of the present investigation it is important to consider whether there is any possibility that the true f-value is actually several orders smaller. This might be the case if (i) the normally accepted electron configuration of either of the two states involved were erroneous or if (ii) there were gross cancellation in the transition integral (as occurs, for example, in that associated with the photo-ionization of atomic potassium (Bates 1947)). (i) seems unlikely on general grounds, and (ii) can probably be excluded, as it would result in the dipole moment being very sensitive to the internuclear distance, which would make calculations on the relative intensities of the individual bands inaccurate if carried out to the usual degree of approximation (based on the Franck-Condon principle), whereas the results of such calculations are, in fact, in fair agreement with observation (Bates 1949).

With an f-value of 0·04 it can be seen from (A. 5) that A, the spontaneous transition probability, is $1{\cdot}7 \times 10^7\,\text{sec.}^{-1}$; the radiative lifetime τ corresponding to this is $6 \times 10^{-8}\,\text{sec.}$* A and τ are not, of course, true constants of the system (they are functions of the p_{ik}'s and λ_{ik}'s of the individual bands); in using the equations given at the beginning of this appendix and taking the wave-length λ appearing in them to be 3914 Å, it is implicitly assumed that the vibrational energy is negligible compared with the electronic energy which is not strictly the case. The related quantity G (defined in the main text) does not, however, depend on this approximation. Its value is given simply by the formula

$$G = \{\lambda_{ik}^3 A_{ik}/c^3 p_{ik}\} \times 10^{-24}\,\text{sec.}^2 \tag{A. 16}$$

$$= \{\lambda^3 A/c^3\} \times 10^{-24}\,\text{sec.}^2, \tag{A. 17}$$

both λ_{ik} and λ being in Ångström units. Numerical substitution at once gives G to be $3{\cdot}8 \times 10^{-38}\,\text{sec.}^2$.

REFERENCES

Allen, C. W. 1946 *Terr. Magn. Atmos. Elect.* **51**, 1.
Alpert, J. L. & Einberg, A. A. 1947 *Bull. Acad. Sci.* (*sér. géogr. géophys.*) U.R.S.S. **11**, 137.
Appleton, E. V. 1937 *Proc. Roy. Soc.* A, **162**, 451.
Appleton, E. V. & Naismith, R. 1935 *Proc. Roy. Soc.* A, **150**, 685.
Appleton, E. V. & Naismith, R. 1939 *Phil. Mag.* **27**, 144.
Appleton, E. V. & Naismith, R. 1940 *Proc. Phys. Soc.* **52**, 402.
Babcock, H. W. & Johnson, J. J. 1941 *Astrophys. J.* **94**, 271.
Barbier, D. 1947 *Ann. Astrophys.* **10**, 141.
Bates, D. R. 1939 *Mon. Not. R. Astr. Soc.* **100**, 25.
Bates, D. R. 1946*a* *Mon. Not. R. Astr. Soc.* **106**, 432.
Bates, D. R. 1946*b* *Mon. Not. R. Astr. Soc.* **106**, 509.
Bates, D. R. 1947 *Proc. Roy. Soc.* A, **188**, 350.
Bates, D. R. 1948 *Emmission spectra of night sky and aurorae* (*Gassiot Conference*), London, Phys. Soc. p. 21.
Bates, D. R. 1949 *Proc. Roy. Soc.* A, **196**, 217.
Bates, D. R. & Massey, H. S. W. 1946 *Proc. Roy. Soc.* A, **187**, 261.

---

* In an earlier paper Bates (1948) took τ to be $10^{-7}\,\text{sec.}$ This slightly higher value was an arbitrary estimate.

Bates, D. R. & Massey, H. S. W. 1947 *Proc. Roy. Soc.* A, **192**, 1.

Berkner, L. V. 1939 *Terrestrial magnetism and electricity* (ed. J. A. Fleming). New York and London: McGraw Hill Book Co.

Bernard, R. 1947 *Ann. Géophys.* **3**, 63.

Bondi, H., Hoyle, F & Lyttleton, R. A. 1947 *Mon. Not. R. Astr. Soc.* **107**, 184.

Bradbury, N. E. 1938 *Terr. Magn. Atmos. Elect.* **43**, 55.

Brown, F. W., Bartlett, J. H. & Dunn, C. G. 1933 *Phys. Rev.* **44**, 296.

Cabannes, J. & Dufay, J. 1946 *Ann. Géophys.* **2**, 290.

Cario, G. 1938 Private communication to J. Gauzit (1938).

Chapman, S. 1931 *Proc. Roy. Soc.* A, **132**, 353.

Coster, D. & Brons, H. H. 1932 *Z. Phys.* **73**, 747.

Denisse, J. F., Seligmann, P. & Gallet, R. 1947 *C.R. Acad. Sci., Paris*, **225**, 1169.

Dufay, J. 1943*a* *Ann. Univ. Lyon* (3rd series).

Dufay, J. 1943*b* *Ann. Astrophys.* **6**, 81.

Dufay, J. & Déjardin, G. 1946 *Ann. Géophys.* **2**, 249.

Dufay, M. & Dufay, J. 1947 *C.R. Acad. Sci., Paris*, **224**, 1834.

Elvey, C. T. 1942 *Rev. Mod. Phys.* **14**, 140.

Elvey, C. T. & Farnsworth, A. 1942 *Astrophys. J.* **96**, 451.

Elvey, C. T., Swings, P. & Linke, W. 1941 *Astrophys. J.* **93**, 337.

Fassbender, M. 1924 *Z. Phys.* **30**, 73.

Gauzit, J. 1938 *Ann. Astrophys.* **1**, 334.

Gaydon, A. G. 1947 *Dissociation energies*. London: Chapman and Hall.

Ghosh, S. N. 1943 *Proc. Nat. Inst. Sci. India*, **9**, 301.

Götz, P. 1947 *Experienta*, **3**, 185.

Harang, L. 1945 *Geofys. Publ. Oslo*, **16**, no. 6.

Herzberg, G. 1939 *Molecular spectra and molecular structure—diatomic molecules*. New York: Prentice Hall.

Herzberg, G. 1942 *J. Chem. Phys.* **10**, 306.

Hoyle, F. 1949 Forthcoming Cambridge Physical Tract.

Hoyle, F. & Bates, D. R. 1948 *Terr. Magn. Atmos. Elect.* **53**, 51.

Hulburt, E. O. 1938 *J. Opt. Soc. Amer.* **28**, 227.

Hulburt, E. O. 1939 *Terrestrial magnetism and electricity* (ed. J. A. Fleming). New York and London: McGraw Hill Book Co.

Hulst, H. C. van de 1948 *Astrophys. J.* **107**, 220.

Hunter, A. 1942 *Rep. Progr. Phys.* **9**, 5.

Johnson, M. 1947 *Mon. Not. R. Astr. Soc.* **107**, 452.

Kiepenheuer, K. O. 1945 *Ann. Astrophys.* **8**, 210.

Martyn, D. F. 1947 *Proc. Roy. Soc.* A, **189**, 241.

Massey, H. S. W. & Bates, D. R. 1943 *Rep. Progr. Phys.* **9**, 62.

Megrelishvili, T. G. 1946 *C.R. Acad. Sci. U.R.S.S.* **53**, 123.

Minnaert, M., Mulders, G. F. W. & Houtgast, J. 1940 *Photometric atlas of the solar spectrum*. Amsterdam: Schnabel, Kampert and Helm.

Mitra, S. K. 1943 *Sci. Cult. India*, **9**, 46.

Mitra, S. K. 1945 Joykissen Maokerjee Medel Lecture, Indian Ass. Cultivation Sci.

Mitra, S. K., Bhar, J. N. & Ghosh, S. N. 1938 *Ind. J. Phys.* **12**, 455.

Mohler, F. L. 1940 *Bur. Stand. J. Res., Wash.*, **25**, 507.

Mulliken, R. S. 1932 *Rev. Mod. Phys.* **4**, 1.

Mulliken, R. S. 1939*a* *J. Chem. Phys.* **7**, 14.

Mulliken, R. S. 1939*b* *J. Chem. Phys.* **7**, 20.

Nicolet, M. 1945 *Mem. R. Met. Inst. Belgium*, **19**, 1.

Papalexi, N. D. 1946 *Bull. Acad. Sci. U.R.S.S.* (ser. phys.), **10**, 237.

Pettit, E. 1940 *Astrophys. J.* **91**, 159.

Roach, F. E. 1939 *Astrophys. J.* **89**, 99.

Robertson, N. G. & Pease, R. N. 1942 *J. Chem. Phys.* **10**, 490.

Rydbeck, O. E. H. 1946 *Chalmers tekn. Högsk. Handl.* no. 53.

Saha, M. N. 1937 *Proc. Roy. Soc.* A, **160**, 155.

Scharonow, W. W. 1944 *C.R. Acad. Sci. U.R.S.S.* **42**, 296.

Schneider, E. G. 1940 *J. Opt. Soc. Amer.* **30**, 128.

Slipher, V. M. 1933 *Mon. Not. R. Astr. Soc.* **93**, 666.

Stebbing, J., Whitford, A. E. & Swings, P. 1945 *Astrophys. J.* **101**, 39.
Störmer, C. 1947 *Observatory*, **67**, 161.
Vegard, L. 1939 *Terrestrial magnetism and electricity* (ed. J. A. Fleming). New York and London: McGraw Hill Book Co.
Waldmeier, M. 1944 *Helv. Phys. Acta*, **17**, 168.
Whipple, F. L. 1943 *Rev. Mod. Phys.* **15**, 246.
White, J. U. 1940 *J. Chem. Phys.* 8, 79, 459.
Woolley, R. v. d. R. 1946 *Proc. Roy. Soc.* A, **187**, 102.
Wulf, O. R. & Deming, L. S. 1936 *Terr. Magn. Atmos. Elect.* **41**, 299.
Wulf, O. R. & Deming, L. S. 1938 *Terr. Magn. Atmos. Elect.* **43**, 283.

INDEX TO VOLUME 196 (A)

Acetaldehyde polymerization (Bevington & Norrish), 363.
Astbury, N. F. Some theoretical considerations on the dynamic properties of plastics, 92.

Bacon, G. E. & Thewlis, J. Neutron diffraction, 50.
Bakerian Lecture. Nuclear explosions (Thomson), 311.
Bates, D. R. The emission of the negative system of nitrogen from the upper atmosphere and the significance of the twilight flash in the theory of the ionosphere, 562.
Bates, D. R. The intensity of distribution in the nitrogen band systems emitted from the earth's upper atmosphere, 217.
Bevington, J. C. & Norrish, R. G. W. The polymerization of acetaldehyde at low temperatures, 363.
Bragg, Sir Lawrence & Lomer, W. M. A dynamical model of a crystal structure. II, 171.
Bretscher, E., Cook, G. B., Martin, G. R. & Wilkinson, D. H. Preparation and properties of a reproducible neutron standard, 436.

Carbon monoxide and carbon dioxide, electronic structures (Moffitt), 524.
Chalmers, B. Some crystal-boundary phenomena in metals, 64.
Cherry, T. M. Flow of a compressible fluid about a cylinder. II. Flow with circulation, 1.
Cherry, T. M. Numerical solutions for transonic flow, 32.
Coles, B. R., Hume-Rothery, W. & Myers, H. P. The structure and properties of the alloy Cu_2MnIn, 125.
Cook, G. B. *See* Bretscher, Cook, Martin & Wilkinson.
Cosmic rays producing nuclear disintegrations (Harding, Lattimore & Perkins), 325.
Couteur, K. J. Le. The indefinite metric in relativistic quantum mechanics, 251.
Crystal-boundary phenomena in metals (Chalmers), 64.
Crystal structure, model (Bragg & Lomer), 171; (Lomer), 182.
Cullis, C. R., Hinshelwood, Sir Cyril & Mulcahy, M. F. R. Influence of chlorine substitution on the rate of oxidation of paraffin hydrocarbons, 160.
Cu_2MnIn alloy, structure and properties (Coles, Hume-Rothery & Myers), 125.

Daly, E. F. The spectral response of a phosphor to infra-red radiation, 554.
Domb, C. Order-disorder statistics. I, 36.
Dynamical model of a crystal structure. II (Bragg & Lomer), 171.
Dynamical model of a crystal structure. III (Lomer), 182.

Electronic structures of carbon monoxide and carbon dioxide (Moffitt), 524.
Emission of the negative system of nitrogen from the upper atmosphere and the significance of the twilight flash in the theory of the ionosphere (Bates), 562.
Experimental and theoretical studies of the behaviour of slow electrons in air. I (Huxley & Zaazou), 402.
Experimental and theoretical studies of the behaviour of slow electrons in air. II. Ionospheric and other applications (Huxley), 427.

Flame propagation, influence of diffusion (Gaydon & Wolfhard), 105.
Flow of a compressible fluid about a cylinder. II. Flow with circulation (Cherry), 1.

Gaydon, A. G. & Wolfhard, H. G. The influence of diffusion on flame propagation, 105.
General kinetic theory of liquids. VI. The equation of state (Rodriguez), 73.
Gouy diffusiometer; further calibration (Ogston), 272.

Harding, J. B., Lattimore, S. & Perkins, D. H. Nuclear disintegrations produced by cosmic rays, 325.
Hartree field and molecular orbitals (Moffitt), 510.
Hinshelwood, Sir Cyril. *See* Cullis, Hinshelwood & Mulcahy.
Hume-Rothery, W. *See* Coles, Hume-Rothery & Myers.
Huxley, L. G. H. Experimental and theoretical studies of the behaviour of slow electrons in air. II. Ionospheric and other applications, 427.
Huxley, L. H. G. & Zaazou, A. A. Experimental and theoretical studies of the behaviour of slow electrons in air. I, 402.
Hydrocarbon molecule, vibration spectra (Sheppard & Sutherland), 195.
Hydrodynamics, moving axes (McVittie), 285.

Indefinite metric in relativistic quantum mechanics (Le Couteur), 251.
Influence of chlorine substitution on the rate of oxidation of paraffin hydrocarbons (Cullis, Hinshelwood & Mulcahy), 160.
Influence of diffusion on flame propagation (Gaydon & Wolfhard), 105.
Intensity distribution in the nitrogen band systems emitted from the earth's upper atmosphere (Bates), 217.

Kinetics of certain reactions between methyl halides and anions in water (Moelwyn-Hughes), 540.
Kinetics of the interaction of atomic hydrogen with olefines (Melville & Robb):
  I. Apparatus and use of para-hydrogen techniques, 445.
  II. Diffusion theory, 466.
  III. Theory and technique involved in the use of the oxides of molybdenum and tungsten as hydrogen atom removers, 479.
  IV. Results obtained by the foregoing techniques, 494.
Kolsky, H., Lewis, J. P., Sampson, M. T., Shearman, A. C. & Snow, C. I. Splashes from underwater explosions, 379.

Lambert, J. D., Roberts, G. A. H., Rowlinson, J. S. & Wilkinson, V. J. The second virial coefficients of organic vapours, 113.
Lattimore, S. *See* Harding, Lattimore & Perkins.
Lewis, J. P. *See* Kolsky, Lewis, Sampson, Shearman & Snow.
Liquids, kinetic theory (Rodriguez), 73.
Lomer, W. M. A dynamical model of a crystal structure. III, 182.
Lomer, W. M. *See* Bragg & Lomer.
Low temperature polymerization of acetaldehyde (Bevington & Norrish), 363.

McVittie, G. C. A systematic treatment of moving axes in hydrodynamics, 285.
McVittie, G. C. Two-dimensional fluid motion referred to a network of orthogonal curves, 301.
Martin, G. R. *See* Bretscher, Cook, Martin & Wilkinson.
Massey, H. S. W. *See* Tsi-Ming Hu & Massey.
Melville, H. W. & Robb, J. C. The kinetics of the interaction of atomic hydrogen with olefines:
  I. Apparatus and use of para-hydrogen techniques, 445.
  II. Diffusion theory, 466.
  III. Theory and technique involved in the use of the oxides of molybdenum and tungsten as hydrogen atom removers, 479.
  IV. Results obtained by the foregoing techniques, 494.
Methyl halides and anions, reactions (Moelwyn-Hughes), 540.
Moelwyn-Hughes, E. A. The kinetics of certain reactions between methyl halides and anions in water, 540.

Moffitt, W. Molecular orbitals and the Hartree field, 510.
Moffitt, W. The electronic structures of carbon monoxide and carbon dioxide, 524.
Molecular orbitals and the Hartree field (Moffitt), 510.
Mulcahy, M. F. R. *See* Cullis, Hinshelwood & Mulcahy.
Myers, H. P. *See* Coles, Hume-Rothery & Myers.

Neutron and proton, interactions (Tsi-Ming Hu & Massey), 135.
Neutron diffraction (Bacon & Thewlis), 50.
Neutron standard (Bretscher *et al.*), 436.
Nitrogen band systems from earth's upper atmosphere (Bates), 217.
Non-central interactions between neutron and proton (Tsi-Ming Hu & Massey), 135.
Norrish, R. G. W. *See* Bevington & Norrish.
Nuclear disintegrations produced by cosmic rays (Harding, Lattimore & Perkins), 325.
Nuclear explosions. Bakerian Lecture (Thomson), 311.
Numerical solutions for transonic flow (Cherry), 32.

Ogston, A. G. The Gouy diffusiometer; further calibration, 272.
Order-disorder statistics. I (Domb), 36.

Paraffin hydrocarbons, rate of oxidation (Cullis, Hinshelwood & Mulcahy), 160.
Pauling, L. A resonating-valence-bond theory of metals and intermetallic compounds, 343.
Perkins, D. H. *See* Harding, Lattimore & Perkins.
Plastics, dynamic properties (Astbury), 92.
Polymerization of acetaldehyde at low temperature (Bevington & Norrish), 363
Preparation and properties of a reproducible neutron standard (Bretscher, Cook, Martin & Wilkinson), 436.
Presidential Address (Robinson), v.

Resonating-valence-bond theory of metals and intermetallic compounds (Pauling), 343.
Robb, J. C. *See* Melville & Robb.
Roberts, G. A. H. *See* Lambert, Roberts, Rowlinson & Wilkinson.
Robinson, Sir Robert. Presidential Address, v.
Rodriguez, A. E. A general kinetic theory of liquids. VI. The equation of state, 73.
Rowlinson, J. S. *See* Lambert, Roberts, Rowlinson & Wilkinson.

Sampson, M. T. *See* Kolsky, Lewis, Sampson, Shearman & Snow.
Second virial coefficients of organic vapours (Lambert, Roberts, Rowlinson & Wilkinson), 113.
Shearman, A. C. *See* Kolsky, Lewis, Sampson, Shearman & Snow.
Sheppard, & Sutherland, G. B. B. M. Vibration spectra of hydrocarbon molecules. I. Frequencies due to deformation vibrations of hydrogen atoms attached to a double bond, 195.
Slow electrons in air, studies (Huxley & Zaazou), 402.
Some crystal-boundary phenomena in metals (Chalmers), 64.
Some theoretical considerations on the dynamic properties of plastics (Astbury), 92.
Snow, C. I. *See* Kolsky, Lewis, Sampson, Shearman & Snow.
Spectral response of a phosphor to infra-red radiation (Daly), 554.
Splashes from underwater explosions (Kolsky, Lewis, Sampson, Shearman & Snow), 379.
Structure and properties of the alloy Cu_2MnIn (Coles, Hume-Rothery & Myers), 125.
Sutherland, G. B. B. M. *See* Sheppard & Sutherland.
Systematic treatment of moving axes in hydrodynamics (McVittie), 285.

Thomson, Sir George, Nuclear explosions. Bakerian Lecture, 311.
Thewlis, J. *See* Bacon & Thewlis.

Transonic flow, numerical solutions (Cherry), 32.
Tsi-Ming Hu & Massey, H. S. W. Non-central interactions between neutron and proton, 135.
Twilight flash in the theory of the ionosphere (Bates), 562.
Two-dimensional fluid motion referred to a network of orthogonal curves (McVittie), 301.

Underwater explosions (Kolsky *et al.*), 379.

Vibration spectra of hydrocarbon molecules. I. Frequencies due to deformation vibrations of hydrogen atoms attached to a double bond (Sheppard & Sutherland), 195.

Wilkinson, D. H. *See* Bretscher, Cook, Martin & Wilkinson.
Wilkinson, V. J. *See* Lambert, Roberts, Rowlinson & Wilkinson.
Wolfhard, H. G. *See* Gaydon & Wolfhard.

Zaazou, A. A. *See* Huxley & Zaazou, A. A.

I. A. R. I. 75

INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY,
NEW DELHI.

Date of issue.	Date of issue.	Date of issue.
9.4.57		
13.2.62		

MGIPC—S5—38 AR/54—7-7-54—7,000.